```
OM nucleic - nucleic search, using sw model

            Copyright (c) 1993 - 2005 Compugen Ltd.
            GenCore version 5.1.6

Run on:     May  8, 2005, 22:38:28 ; Search time 1549 Seconds
                                    (without alignments)
                                    10618.584 Million cell updates/sec

Title:          US-10-751-550-1
Perfect score:  2690
Sequence:       1 tctagataatacgaccact.........ccataactagtaccatgg 2690

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       5654200 seqs, 3057283753 residues

Total number of hits satisfying chosen parameters:    11308400

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      Published_Applications_NA:*
       1:  /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
       2:  /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
       3:  /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
       4:  /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
       5:  /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
       6:  /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
       7:  /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
       8:  /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
       9:  /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
      10:  /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
      11:  /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
      12:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
      13:  /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
      14:  /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
      15:  /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
      16:  /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
      17:  /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
      18:  /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
      19:  /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
      20:  /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
      21:  /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
      22:  /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2690	100.0	2690	19	US-10-751-550-1	Sequence 1, Appli
2	843.8	31.4	987	18	US-10-845-059-10	Sequence 10, Appl
3	192.4	7.2	4830	18	US-10-437-961-1638	Sequence 1638, Ap
4	160.4	6.0	4920	18	US-10-425-115-54556	Sequence 54556, A
5	160.4	6.0	7593	18	US-10-425-115-159798	Sequence 159798,
6	158.8	5.9	6624	18	US-10-425-115-54546	Sequence 54546, A
7	157.2	5.8	3975	18	US-10-425-115-54557	Sequence 54557, A
8	156.8	5.8	1031	18	US-10-437-963-4747	Sequence 4747, Ap
9	156.2	5.8	1126	18	US-10-437-963-85940	Sequence 85940, A
10	155.6	5.8	5913	18	US-10-425-115-159738	Sequence 159738,
11	155	5.8	4014	18	US-10-437-963-25937	Sequence 25937, A

c	No.	Score		Length	DB	ID	Description
c	12	155	5.8	7495	18	US-10-437-963-26057	Sequence 26057, A
	13	154.6	5.7	5928	18	US-10-437-963-85935	Sequence 85935, A
	14	154	5.7	5460	18	US-10-425-115-54555	Sequence 54555, A
	15	154	5.7	7170	18	US-10-425-115-159747	Sequence 159747, A
	16	153.4	5.7	4884	18	US-10-437-963-17763	Sequence 17763, A
	17	152.4	5.7	3792	18	US-10-425-115-159762	Sequence 159762,
	18	152.4	5.7	4017	18	US-10-425-115-159761	Sequence 159761,
	19	151.8	5.6	6108	18	US-10-437-963-26000	Sequence 26000, A
	20	151.4	5.6	2000	17	US-10-260-238-2549	Sequence 2549, Ap
	21	150.8	5.6	5547	18	US-10-425-115-54553	Sequence 54553, A
	22	150.2	5.6	4398	18	US-10-437-963-25899	Sequence 25899, A
	23	150.2	5.6	5700	18	US-10-437-963-25816	Sequence 25816, A
	24	149.2	5.5	4797	17	US-10-260-238-1495	Sequence 1495, Ap
	25	149.2	5.5	4989	18	US-10-437-963-25811	Sequence 25811, A
	26	148.8	5.5	5548	18	US-10-437-963-25978	Sequence 25978, A
	27	148.6	5.5	4425	18	US-10-437-963-26103	Sequence 26103, A
	28	148.6	5.5	4434	18	US-10-437-963-26124	Sequence 26124, A
	29	147	5.5	4860	18	US-10-437-963-26015	Sequence 26015, A
	30	147	5.5	6015	18	US-10-437-963-25902	Sequence 25902, A
	31	146	5.4	3144	18	US-10-425-115-159765	Sequence 159765,
c	32	146	5.4	4272	18	US-10-425-115-71473	Sequence 71473, A
	33	146	5.4	5901	18	US-10-425-115-159736	Sequence 159736,
	34	145.8	5.4	6686	18	US-10-425-115-54545	Sequence 54545, A
	35	145.4	5.4	3615	18	US-10-437-963-25886	Sequence 25886, A
	36	144.2	5.4	1089	18	US-10-425-115-166297	Sequence 166297,
	37	143.8	5.3	918	18	US-10-437-963-25867	Sequence 25867, A
	38	142.6	5.3	4809	18	US-10-425-115-159746	Sequence 159746,
	39	140.8	5.2	5529	18	US-10-437-963-26012	Sequence 26012, A
	40	139	5.2	7567	18	US-10-437-963-26107	Sequence 26107, A
c	41	136.6	5.1	4557	18	US-10-437-963-26006	Sequence 26006, A
	42	136.6	5.1	6209	18	US-10-425-115-159776	Sequence 159776,
	43	134.2	5.0	4755	18	US-10-437-963-25893	Sequence 25893, A
	44	133.6	5.0	5523	18	US-10-437-963-25999	Sequence 25999, A
	45	132.6	4.9	5157	18	US-10-437-963-25829	Sequence 25829, A

ALIGNMENTS

```
RESULT 1
US-10-751-550-1
; Sequence 1, Application US/10751550
; Publication No. US20050034192A1
; GENERAL INFORMATION:
;  APPLICANT: The Texas A&M University System
;  TITLE OF INVENTION: STEM-REGULATED, PLANT DEFENSE PROMOTER
;  TITLE OF INVENTION:  AND USES THEREOF IN TISSUE-SPECIFIC EXPRESSION IN MONOCOTS
;  FILE REFERENCE: 017575.0775
;  CURRENT APPLICATION NUMBER: US/10/751,550
;  CURRENT FILING DATE:  2004-01-05
;  PRIOR APPLICATION NUMBER: US 60/437,974
;  PRIOR FILING DATE: 2003-01-03
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;  LENGTH: 2690
;  TYPE: DNA
;  ORGANISM: Sugarcane
;  FEATURE:
;  NAME/KEY: promoter
;  LOCATION: (1)...(2686)
US-10-751-550-1

Query Match             100.0%;  Score 2690;  DB 19;  Length 2690;
Best Local Similarity   100.0%;  
Matches 2690;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy          1 TCTAGATAATACGACTCACTATAGGGCACGCGTGGTCGACGGCCCGGGCTGGTCGTCGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TCTAGATAATACGACTCACTATAGGGCACGCGTGGTCGACGGCCCGGGCTGGTCGTCGAC 60

Qy         61 AGCTAGAGGCGCCACCGCGGTCCTAGCTTCCTCCAACTTCTCGTCGGAGATCCCTTCAGGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

(This page is a rotated pairwise nucleotide sequence alignment between two sequences labelled "Qy" (query) and "Db" (database), shown with position numbering along both margins. The alignment spans residues 61–2280.)

Qy 2281 TGCTCGGAGACAAGCAATTGTGGCCAGTCACTATAGCTACGTCAGAGGGTAGTGGGAGCA 2340
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2281 TGCTCGGAGACAAGCAATTGTGGCCAGTCACTATAGCTACGTCAGAGGGTAGTGGGAGCA 2340

Qy 2341 GTTGCGTCGTTGATTGAAAACAGGTGATCGTATCAGATATTGCATTCACATGGACA 2400
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2341 GTTGCGTCCTTGATTGAATTGAAAACAGGTGATCGTATCAGATATTGCATTCACATGGACA 2400

Qy 2401 GTAAATGTGGTACAGTAACTTCGCAAACAATAAAATCTGTCACAATTTATTAGTGCACTC 2460
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2401 GTAAATGTGGTACAGTAACTTCGCAAACAATAAAATCTGTCACAATTTATTAGTGCACTC 2460

Qy 2461 CTCTGACGTAAATGCTTCTACGTCAGAGGATTTGATTCCGAGGGCCGCTGCACCCATCAC 2520
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2461 CTCTGACGTAAATGCTTCTACGTCAGAGGATTTGATTCCGAGGGCCGCTGCACCCATCAC 2520

Qy 2521 TAATGACGGTCTTTACCCATCATCATGGACCATTGTTCACATCCATGTCATCACTGTCGT 2580
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2521 TAATGACGGTCTTTACCCATCATCATGGACCATTGTTCACATCCATGTCATCACTGTCGT 2580

Qy 2581 CCTGTCCATGCACTGCAGCCCTCTATAAATACTGGCATCCCTCCCCGTTCACAGATCAC 2640
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2581 CCTGTCCATGCACTGCAGCCCTCTATAAATACTGGCATCCCTCCCCGTTCACAGATCAC 2640

Qy 2641 ACAACACAAGCAAGAAATAAACGGTAGCTGCCATAACTAGTACACCATGG 2690
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2641 ACAACACAAGCAAGAAATAAACGGTAGCTGCCATAACTAGTACACCATGG 2690

RESULT 2
US-10-845-059-10
; Sequence 10, Application US/10845059
; Publication No. US2005005323A1
; GENERAL INFORMATION:
; APPLICANT: Birch, Robert George
; APPLICANT: Wu, Luguang
; TITLE OF INVENTION: METHOD FOR INCREASING PRODUCT YIELD
; FILE REFERENCE: 900145.401
; CURRENT APPLICATION NUMBER: US/10/845,059
; CURRNT FILING DATE: 2004-05-12
; PRIOR APPLICATION NUMBER: AU2003902253
; PRIOR FILING DATE: 2003-05-12
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; LENGTH: 987
; TYPE: DNA
; ORGANISM: Saccharum sp.
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(987)
; OTHER INFORMATION: DNA sequence of promoter P67B
US-10-845-059-10

Query Match   31.4%;   Score 843.8;   DB 18;   Length 987;
Best Local Similarity   91.7%;   Pred. No. 6.3e-220;
Matches 947;   Conservative 0;   Mismatches 32;   Indels 54;   Gaps 3;

Qy 1639 GAGCTCTCAATGGGAGGTGCTCGAAGACATATTACCCATGTGTATGGCAAGATGTTTAGC 1698
        ||||||| |||||||||||||||||||||||||| ||| |||||||||||||||||||||
Db   1 GAGCTCTCGATGGGAGGTGCTCGAAGACATATTAGCCAAGTGTATGGCAAGATGTTTAGC 60

Qy 1699 TAGTAACTGACTGATAGTGTAAACGATCTCCAATGGGCAAGACATATTACCTAAGGCCA 1758
        ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TAGTAGCTGACTGATAGTGTAAACGATCTCCAATGGGCAAGACATATTACCTAAGGCCA 120

Qy 1759 GGCTGGTTTTTGCAAGTTGCAAGTTGCAAGTTGTGCGGAGTTGTAAACGATCT 1818
        ||||||||||||||||||| |||||||||||||||||||||
Db 121 GGCTGGTTTTTGCAAGTTGAGTTGAGTTGTGCGGAGTTGTAAACGATCT 180

Qy 1819 CCAATGGGCAAGACATCCTACCCTATATATAGTGAAGGGGCAGTAGCTGATTGAGAATC 1878
        ||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db 181 CCAATGGGCAAGACATCCTCAACCTATATATAGTGAAGGGGCAGTAGCTGATTGAGAATC 240

Qy 1879 AATCAATCAAGCACAATATAATTTATTAATTTTTATATTCAAACCCAATTTTTTTTCCTTTT 1938
        | |||||||||||| ||||||||||||||||| ||| |||||||||||||| |||||||||
Db 241 AACCAATCAAGCACACATATAATTTATTAATTTTTTATTAATTTTTATTCAAACCCAA-TTTTTTCCTTTT- 299

Qy 1939 CCAACCCTAATTATAGTTTCCTTTGCCTTTGCCTCTAGGACAAATTGACGTGTTCCGGGTATCC 1998
        |||||||||||||||||||| ||||| |||||||||||||||||||||||||| |||||||||
Db 300 CCAACCCTAATTATAGTTTCTCCTTTGCCTCTAGGACAAATTGACGTGTTCCTGGTATC- 358

Qy 1999 TGCTGAATTAAGAACAACCCTAGGTGCACCTGTCCCGATAGAGTCCCACCTGGGTAGGCA 2058
                                               ||||||||||||
Db 359 ------------------------------------CCTGGGTAGGCA 370

Qy 2059 TTCATAGGGATTCGTGTATTCCTGCAAAAAAGCCGATTAAGCTGGCTTCTAAAACTGGCT 2118
        |||||||||| | ||||| |||||||||||||||||||||||||||||||||||||||||
Db 371 TTCATAGGGATTACGGGTATTTCCTGCAAAAAAGCCGATTAAGCTGGCTTCTAAAACTGGCT 430

Qy 2119 AGGGCCGGATTCGTGTGGCCTTCACATACCAGTGATTTCATGTGATCCGTGCATTCTAGCA 2178
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 431 AGGGCCGGATTCGTGTGGCCTTCACATACCAGTGATTTCATGTGATCCGTGCATTCTAGCA 490

Qy 2179 CTTTGCTATGTAACCCAAACTTAAGTGACAACTATAAAATGCTACTTGCAGGATGTTA 2238
        ||||||| |||||||||||| ||||||||||||||||||||||||||||||||
Db 491 CTTTGCTGTGTAACCCAAACTGATGTCGACAACTATAAAATGCTACTTGCAGGATGTTA 550

Qy 2239 TCAGACACAACTCCTAATCTACGGAAGCCTAAGTTAGTTTGCTCGGAGACAAGCAAT 2298
        ||| |||||||||| |||||||||||||||||| ||||||||||||||||||||||
Db 551 TCATGACACAACTCCCTAAATCTACGGAAGCCTAAGTTTAGTTTGCTCGGAGACAAGCAAT 610

Qy 2299 TGTGGCCAGTCACTATAGCTACGTCAGAGGGTAGTGGGAGCAGTTGCGTCGTTGATTGA 2358
        |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db 611 TGTGGCCAGTCACTTTAGCTACGTCAGAGGGTAGTGGGAGCAGTTGCGTCGTTGATTGA 670

Qy 2359 AAACAGGTGGATCGTATCAGATATTCACATGGACAGCAGTAAATGTGGTACAGTAA 2418
        ||||||||| ||| || |||||||| ||||||||||||||||||||||||||
Db 671 AAACAGGTGGATCATCATTAGATATT---ATTCACATGGACAGCAGTAAATGTGGTACAGTAA 726

Qy 2419 CTTCGCAAACAATAAAATCTGTCACAATTTATTAGTGCACTCCTCTGACGTAAATGCTTC 2478
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 727 CTTCGCAAACAATAAAATCTGTCACAATTTATTAGTGCACTCCTCTGACGTAAATGCTTC 786

Qy 2479 TACGTCAGAGGATTGATTCCGAGGGCCGCTGCACCCATCACTAATGACGGTCTTTACCC 2538
        ||||||||||||| || |||||||| || ||||||||||||||||||||||||||||||
Db 787 TACGTCAGAGGATTTGAGTCCGAGGGTGCGCACCCATCACTAATGACGGTCTTTACCC 846

Qy 2539 ATCATCATGGACCATTGTTCACATCCATGTCATCACTCGTCGTCCATGCACTGCAG 2598
        ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db 847 ATCATCATGGACCATTGTTCACATCCATGTCGTCGTCCATGCACTGCAG 906

Qy 2599 CCCTCTATAAATACTGGCATCCCTCCCCGTTCACAGATCACACACAAGCAAGAAAT 2658
        ||||||||||||||||||| |||||||||||||||||||||| ||||||||||||
Db 907 CCCTCTATAAATACTGGCACCCCTCCCCGTTCACAGATCACACCACAAGCAAGAAAT 966

Qy 2659 AAACGGTAGCTGC 2671
        |||||||||||||
Db 967 AAACGGTAGCTGC 979

RESULT 3
US-10-437-963-1638
; Sequence 1638, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 1638
; LENGTH: 4830
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_101485C.1
US-10-437-963-1638

Query Match   7.2%; Score 192.4; DB 18; Length 4830;
Best Local Similarity 67.2%; Pred. No. 5.2e-41;
Matches 328; Conservative 0; Mismatches 136; Indels 24; Gaps 3;

```
Qy  1232 GAGCTCATCATTTATATAAGCATGTGATACCAAAATTACTCCTTTGAGCATGTTTATA 1291
         ||||||| ||| ||||||| ||| |||||||||||||||||||||||||||||| |||
Db  4343 GGGCTCATCATCGTATATCTAAGCATGTGCACACAAGGTAACACCATTTGAATTAGTATATG 4402

Qy  1292 TTTAGGAGACGTTTTACCTGTGAGGTAAATCGGCTAAATCGGCTAAGCAAA 1351
Db  4403 GGCAAGAGAGCCCATTTGCAGTGTGAGGTGAGCGCAGCGCCCTTAGATTGGCCAGACAAA 4462

Qy  1352 ATAATTTATCACCTGTGATTCTAACAATTGATGATGAGGACAATATTGATGAGGTGACTG 1411
Db  4463 ATAATTTGTCGGCAGTAGATTATCACAACTTAATGATGGAATAGATGATGAGGTTAGTG 4522

Qy  1412 ACAAATGATTGAAGGCTTTAAAGGAGGATTGAGAAGGATAAAT--------------CT 1455
Db  4523 ATGAAAGATTGAAAGCTTTGAGAGAGATTGAGAAGAATTGAGAGTAGCTAAAGCTT 4582

Qy  1456 ACAATAAAAAAATGTAAAGAAGAAAGCATTCAAAGTGTGAGATCTCGGTGTGTGGAAGACTATTT 1515
Db  4583 ACAATAAAAAAGGTGAAGAAAATCGTTTCAAATTGGAGATTTAAATTGTGTGGAAAACAATTC 4642

Qy  1516 TGCCTCCTTGGGGTAAAGACAACAACAGTTTAGTAAGTGGCTCAAAATTGGGAGGGCCCA 1575
Db  4643 TGCCTGTT-GGTTCAAAAGATAATAAATTTGGGAAAAGGTCTCCAAATTGGGGAGGGGCCA 4701

Qy  1576 TGCAAGATTGTTAAAGTAAT-------TGTTTGGATTGACGGAGGCATTTCAAGGTGAT 1628
Db  4702 TATAGAATTGTAGAAAATAGTCCCCGGAATTCTTATTTTGTGCAAAGTCTACGAGGATAT 4761

Qy  1629 CATCTACCTAGAGCTCTCAATGGGAGGTGCTCGAGAACATATTACCCATGTGTATGGCAA 1688
Db  4762 AAGTTGCCTAGAGCTTTAATGGGAGAATATTTAAAAAGGTACTATCCCAAGTGTTTGGCAA 4821

Qy  1689 GATGTTTA 1696
Db  4822 GATGCTTA 4829
```

RESULT 4
US-10-425-115-54556
; Sequence 54556, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
SEQ ID NO 54556
; LENGTH: 4920
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_149756C.1
US-10-425-115-54556

Query Match   6.0%; Score 160.4; DB 18; Length 4920;
Best Local Similarity 63.1%; Pred. No. 3.1e-32;
Matches 308; Conservative 0; Mismatches 156; Indels 24; Gaps 3;

```
Qy  1232 GAGCTCATCATTTATATAAGCATGTGATACCAAAATTACTCCTTTGAGCATGTTTATA 1291
         ||||||| ||| ||||||| ||| ||||||||||||||||||||||||||||| |||
Db  4433 GGGCTCATCAGATAATCTCAAACATAGTGCTACTAAAGTATCTCCTTTTGAGCTGTCTATG 4492

Qy  1292 TTTAGGAGACGTTTTACCTGTGAGGTAAATCGGCTAAGCAAA 1351
Db  4493 GGCAGGAAGCAGTGTTACCTGTGGAAATAGTTTGAATGCTGTCAGGTTTGCCAGACAAA 4552

Qy  1352 ATAATTTATCACCTGTGATTCTAACAATTTGATGATGGACAATATTGATGAGGTGACTG 1411
Db  4553 ATGATCTAACTGCTACTGATTATTATAATTCAATGATGAATATATTGATGAGGTGACCG 4612

Qy  1412 ACAAATGATTGAAGGCTTTAAAGGAGGATTGAGAAGGATAA---------------ATCT 1455
Db  4613 ACAAAAGGATGATAGCTTTGGGAGCAATAGAAAAGGACAAGATCATAGTAGCCAGGCCT 4672

Qy  1456 ACAATAAAAATGTAAAGAAGAAAGCATTCAAAGTGTGAGATCTCGGTGTGGAAGACTATTT 1515
Db  4673 ACAACAAGAGAGGTCAAAGCAAAATCATTTCAAGTAGGAGGACTTGGTGTGGAAGACCATTC 4732

Qy  1516 TGCCTCCTTGGGGTAAAGACAACAACAGTTTAGTAAGTGGCTCAAAATTGGGAGGGCCCA 1575
Db  4733 TGCCTTCTAAGGAAT-AAAGACCGAAAGTTCGGGAAATGGTCGCCAAGCTGGGAGGGTCCT 4791

Qy  1576 TGCAAGATTGTTAAAGTAATTGTTTTGAT 1628
Db  4792 TATAAAGTGAAACAGGTGATGTCTGTAATGCCTATTTGCTACAAACATTACAAGGCGAA 4851

Qy  1629 CATCTACCTAGAGCTCTCAATGGGAGGTGCTCGAGACATATTACCCATGTGTATGGCAA 1688
Db  4852 GATTTGCCTAAGGCTTGAATGAGCGTTTCCTCAAGCAGTATCATCCCAGTATGTGGCAA 4911

Qy  1689 GATGTTTA 1696
Db  4912 AATGCTTA 4919
```

RESULT 5
US-10-425-115-159798
; Sequence 159798, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
SEQ ID NO 159798
; LENGTH: 7593
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_77309C.1
US-10-425-115-159798

Query Match   6.0%; Score 160.4; DB 18; Length 7593;
Best Local Similarity 63.1%; Pred. No. 4e-32;
Matches 308; Conservative 0; Mismatches 156; Indels 24; Gaps 3;

```
Qy  1232 GAGCTCATCATTTATATAAGCATGTGATACCAAAATTACTCCTTTGAGCATGTTTATA 1291
```

Db 7106 GGGCTCATCATGAATATCTAAGCATGTGCTACTAAGGTATCTCCTTTTGAGCTTGTCTATG 7165
Qy 1292 TTTAGGAGACGTTTTACCTGTTGAGGTAAATCGGCTAATAAATCGGCTAAGCAAA 1351
         ||||||| || |||| ||| ||||||||| ||| ||| ||||||||||
Db 7166 GGCAGGAGCAGCAGTGTTACCTGTGAATGGTCGTGTCAGGTTTGCCAGACAAA 7225

Qy 1352 ATAATTTATCACCTGTGATCTTCAACAATTGATGATGGACAATATTGATGAGGTGACTG 1411
         |||||||| |||| |||||| || |||| |||||||||||||||||||||||
Db 7226 ATGATCTAACTACTACTGATTATTATAATTCAATGATGGACAATATTGATGAGGTGACCG 7285

Qy 1412 ACAAATGATTGAAGGCTTTAAAGGAGATTGAGAAGGATAA------------ATCT 1455
         |||| ||| ||||||| ||||| ||||||||||||||||| ||||
Db 7286 ACAAGAGGATGATAGCTTTGGGAGCAATAGAAAAGGACAAGATCATGGTAGCTAGGGGGCCT 7345

Qy 1456 ACAATAAAAATGTAAAGACAAGCATTCAAAGTCGTGAGATCTGGTGTGGAAGACTATTT 1515
         ||||||||||||||||| |||||||| || ||||| |||||||||||||| |||
Db 7346 ACAACAAGAGGTCAAAGCAAATCATTCCAAGTAGGAGACTGGTGTGGAAGACCATTC 7405

Qy 1516 TGCCTCTTCGGGGTAAAAGACAACAAGTTAGTAAGTGGCCTCAAAATTGGGAGGGCCCA 1575
         ||||| || |||||||||| | || || ||| |||||| || |||||||||||
Db 7406 TGCCTCTAAGGAAT-AAAGACCCGAAAGTCGGCAAATGTGCCAAGCTGGGAGGGGTCCT 7464

Qy 1576 TGCAAGATTGTTAAAGTAATTGTT------TTGGATTGACGGAGGCATTTCAAGGTGAT 1628
         || |||| ||| ||| |||| ||| ||||| ||| | ||| |||| |||
Db 7465 TATAAAGTGAAACAGGTGATGTTTGGTAACGGTTTATTACTACAAACATTACAAGGCAA 7524

Qy 1629 CATCTACCTAGAGCTCTCAATGGGAGGTGCTCGAAGACATATTACCCATGTGTATGGCAA 1688
         | ||||| ||||| ||||||| |||||||| |||| ||| || ||| |||||||||||
Db 7525 GATTTACCTAAGGCTTTGAATGGGCGTTTCCTCAAACAGTACCATCCTAGTATGTGGCAA 7584

Qy 1689 GATGTTTA 1696
         |||| ||
Db 7585 GATGCCTA 7592

RESULT 6
US-10-425-115-54546.
; Sequence 54546, Application US/10425115
; Publication No. US2004021427ZA1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 54546
; LENGTH: 6624
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_149747C.1
US-10-425-115-54546

Query Match   5.9%;
Best Local Similarity   62.9%;   Score 158.8;   DB 18;   Length 6624;
Matches 307;   Conservative   0;   Mismatches 157;   Indels 24;   Gaps 3;   Pred. No. 1e-31;

Qy 1232 GAGCTCATCATTTATATAAGCATGTGATACCAAAATTACTCCTTTTGAGCATGTTTATA 1291
         |||||||| || ||||||||||||| || |||||||||||||||||||| ||| ||
Db 6137 GGGCTCATCATGAATATCTAAGCATGTGCTACTAAGGTATCTCCTTTTGAGCTTGTCTATG 6196

Qy 1292 TTTAGGAGACGTTTTACCTGTTGAGGTAAATCGGCTAAGCAAA 1351
         |||||| || || |||| ||| ||||| ||| ||||||
Db 6197 GGCAGGAGCAGTGTTACCTGTGTCTGAAATAAGTTTGCCTGTCAGTTCGGCCAGACAAA 6256

Qy 1352 ATAATTTATCACCTGTGATTCTAACAATTGATGATGGACAATATTGATGAGGTGACTG 1411
         |||||||| || |||| ||| ||| ||| ||||||||||||||||||||||
Db 6257 ATGATCTAACTACTACTGATTATTATAATTCAATGATGGACAATATTGATGAGGTGACCG 6316

RESULT 7
US-10-425-115-54557
; Sequence 54557, Application US/10425115
; Publication No. US2004021427ZA1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 54557
; LENGTH: 3975
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_149757C.1
US-10-425-115-54557

Query Match   5.8%;
Best Local Similarity   62.7%;   Score 157.2;   DB 18;   Length 3975;
Matches 306;   Conservative   0;   Mismatches 158;   Indels 24;   Gaps 3;   Pred. No. 2e-31;

Qy 1232 GAGCTCATCATTTATATAAGCATGTGATACCAAAATTACTCCTTTTGAGCATGTTTATA 1291
         |||||||| || ||||||||||||| || |||||||||||||||||||| ||| ||
Db 3488 GGGCTCACAGAATATCTAAACATAGTGCTACTAAAGTGTCTCCTTTTGAGCTTGTCTATG 3547

Qy 1292 TTTAGGAGACGTTTTACCTGTTGAGGTAAATCGGCTAAGCAAA 1351
         |||||| || || |||| ||| ||||| ||| ||||||
Db 3548 GGCAGGAGCAGCAGTGTGCCTGTGAAATAAGTTTGCCTGTCAGTTCGGCCAGACAAA 3607

Qy 1352 ATAATTTATCACCTGTTGATTCTAACAATTGATGATGGACAATATTGATGAGGTGACTG 1411
         |||||||| || |||| ||| ||| ||| ||||||||||||||||||||||
Db 3608 ATGATTTAACTCTCCACTGATTTATTACAATGCAATGATGAATATTGATGAGGTGACCG 3667

Qy 1412 ACAAATGATTGAAGGCTTTAAAGGAGATTGAGAAGGATAA------------ATCT 1455
         |||| ||| ||||||| ||||| ||||||||||||||||| ||||
Db 3668 ACAAGAGGGTGATAGCTTTGGGAGGCAATAGAAAAGACAAGATCATCGTAGCCAGGGCCT 3727

Qy 1456 ACAATAAAAATGTAAAGACAAGCATTCAAAGTCGTGAGATCTGGTGTGGAAGACTATTT 1515
         ||||||||||||||||| |||||||| || ||||| |||||||||||||| |||
Db 3728 ACAATAAGAAGGTCAAAGCAAGTCATTTCAGTAGGGGACCTGGTGGTGTGGAAGAAGACCATCC 3787

```
Qy   1516 TGCCCTTGGGGTAAAAGACAACAAGTTTAGTAAGTGGCCTCAAAATTGGAGGGCCCA 1575
          |||||||           ||  |||||||  ||||||||||||||||| |||||||
Db   3788 TGCCCTTCTAAGGAAT-AAAGACCGGAAGTTTGGAAAATGG'CGCCAAGCTGGGAGGGTCCT 3846

Qy   1576 TGCAAGATTGTTAAAGTAATTGTTTGGA------TTGACGGAGGCATTTCAAGGTGAT 1628
          ||||||| |||| ||||  ||||     |          ||||||||  |||||||
Db   3847 TATAAAGTAAAACAGGTAATATCTGGTAACGCCTATTTACTACAAACATTACAGGGCAAG 3906

Qy   1629 CATCTACCTAGAGCTCTCAATGGGAGGTGCTCGAAGACATATTACCCATGTGTATGGCAA 1688
          |||||||||||            || ||||||||||||||||||||||||||||||||
Db   3907 GATTTACCCAAAGCTTGAATGGGCGTTTCCTCAAGCAGTACCACTCCTAGTATGTGGCAA 3966

Qy   1689 GATGTTTA 1696
          |||| |||
Db   3967 GATGCTTA 3974


RESULT 8
US-10-437-963-4747
; Sequence 4747, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 4747
; LENGTH: 1031
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_11601C.1
US-10-437-963-4747

Query Match             5.8%;  Score 156.8;  DB 18;  Length 1031;
Best Local Similarity  62.8%;  Pred. No. 1.1e-31;
Matches  305;  Conservative  0;  Mismatches 157;  Indels 24;  Gaps 3;

Qy   1232 GAGCTCATCATTTATATAAGCATGTGATACCAAAATTACTCCTTTGAGCATGTTTATA 1291
          |||||||||    | ||||||||| ||||  ||  ||| ||| ||||| |||||
Db    547 GAGCTCATAGAATATCTAAGCATGCATGGTGCTAAAATCTGAACGCTAATAAAAATCTATT 606

Qy   1292 TTTAGGAGGACGTTTTACCTGTTGAGGTAAATCTGAACGCTAATAAATCGGCTAAGCAAA 1351
          ||||||||||  ||||| ||| |||||  ||  ||||||| |||| |||||  |||
Db    607 TGCAAGAAGCCGTTTTACTAGTAGAGGTAAATCTTGGCCTCGTGTTAGATATCAAGCAAG 666

Qy   1352 ATAATTTATCACCTGTTGATTCTAACAATTGATGATGACAATATTGATGAGGTGACTG 1411
          || |||||||  ||| ||||| |||||||  ||||||| |||| |||||||||  |||
Db    667 ATGATTTATCGGCCGCCGAAGATATAGGACGTTGATGGGAGCAATCTTGATGATGTGACCG 726

Qy   1412 ACAAATGATTGAAGGCTTTAAAGAAGAAGAATTGAGAAGGATAAA------------TCT 1455
          ||||| |||| ||||||||||||| ||| |||||||  |||||   ||| ||||
Db    727 ATAAACGTCTAAAGGCTTTAGAGAAGGAGAATAGAGAATATAGAGAGATAGAAAAAGCCT 786

Qy   1456 ACAATAAAAATGTAAAGAAGAACAACAAGTTTAGTAAGTGGCCTCAAAATTGGGAGGGCCCA 1515
          ||||||  |||  ||||| ||| ||| ||||||| |||||||||||||||| |||||||||||
Db    787 ATATAAAAAAGAGAGAGAAAAATCATTTCAAGTAGGAAAACTTGGTTTGGAAGACGATTT 846

Qy   1516 TGCCCTTCTAAGGGTAAAAGAACAACAAGTTTAGTAAGTGGCCTCAAAATTGGGAGGGCCCA 1575
          |||||||              |||||||| |||||||||||| |||||||| |||||||
Db    847 TTCCT-TTGGGTTCTAGATGTAAGGACTTCGGTAAGTGGTTGCCTAGAACTTGGTAAGTCCT 905

Qy   1576 TGCAAGATTGTTAAAGTAATTGTTTGGA------TTGACGGAGGCATTTCAAGGTGAT 1628
          ||||||||||||||||||||||||    |           ||||||||||||||||||
Db    906 TATCGGATATGTGGAATCGTTCGAGGGAATCGTGTGTTCTTCGAAGACTTGCAAGGACAA 965
```

```
Db    906 TATCGGATATGTGGAATCGTTCGAGGGAATCGTGTGTTCTTCGAAGACTTGCAAGGAA 965

Qy   1629 CATCTACCTAGAGCTCTCAATGGGAGGTGCTCGAAGACATATTACCCATGTGTATGGCAA 1688
          |||||||||||   ||| ||||||||   |||||||||||||||||||||||||||||
Db   1629 CATCTACCTAGAGCTCTCAATGGGAGGTGCTCGAAGACATATTACCCATGTGTATGGCAA 1688

Qy   1689 GATGTT 1694
          ||||||
Db   1026 GATGCT 1031


RESULT 9
US-10-437-963-85940
; Sequence 85940, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 85940
; LENGTH: 1126
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_85028C.1
US-10-437-963-85940

Query Match             5.8%;  Score 156.2;  DB 18;  Length 1126;
Best Local Similarity  62.6%;  Pred. No. 1.7e-31;
Matches  305;  Conservative  0;  Mismatches 158;  Indels 24;  Gaps 3;

Qy   1232 GAGCTCATCATTTATATAAGCATGTGATACCAAAATTACTCCTTTGAGCATGTTTATA 1291
          |||||||||    | ||||||||| ||||  ||  ||| ||| ||||| |||||
Db    639 GGGCTCATCCGATATCTAAGCATGGTCGACTCAAGTCACCGTATGAGCCTAGTCATG 698

Qy   1292 TTTAGGAGGACGTTTTACCTGTTGAGGTAAATCTGAACGCTAATAAATCGGCTAAGCAAA 1351
          ||||||||||  ||||| ||| |||||  ||  ||||||| |||| |||||||  |||
Db    699 GGCAGAAGGCCGTGTTACTTCTTGTGAGGTTAATTTACAAGTCTTAGAATAGGCCAGACAAA 758

Qy   1352 ATAATTTATCACCTGTTGATTCAACAATTGATGATGACAATATTGATGAGGTGACTG 1411
          || |||||||  || || |||| ||||| || |||| |||| |||||||||||||
Db    759 ATGGTTTGTCCGCTGTAGATTATAGGAATTTAATGGATTTGGCAGGAATATCAAGCAAG 818

Qy   1412 ACAAATGATTGAAGGCTTTAAAGAAGAAGAATTGAGAAGGATAAA------------CT 1455
          | |||||||| ||||||||||| || |||||||||| |||||   || |||
Db    819 AGGAAAGATTGAAGGCCTTGAGACAACATTGAGACAGAAGAAAAAATTAAGAGTAGCTAGGCTT 878

Qy   1456 ACAATAAAAATGTAAAGAAGAACAACAAGTTTAGTAAGTGGCCTCAAAATTGGGAGGGCCCA 1515
          ||||||  |||  ||||| ||| ||| ||||||| ||||||||||||||||| |||||||||||
Db    879 ATAATAAGAGGTGAGAGAAAAATCATTTCAGATAGGTGAGTTGGTATGAGAAAAACCATTTT 938

Qy   1516 TGCCCTTGGGGTAAAAGACAACAACAAGTTTAGTAAGTGGCCTCAAAATTGGGAGGGCCCA 1575
          |||||||       |||||| ||| ||||| |||||||||||| |||||| |||||||||||
Db    939 TGCCCATTGGAACTAGGAG-TAGCAAATTTGGCAGTATTTGGCGGTGGGAAGGTCCT 997

Qy   1576 TGCAAGATTGTTAAAGTAATTGTTTGGATTGA-----?-CGGAGGCATTTCAAGGTGAT 1628
          ||||||||||| |||||||| |||||||    |        | |||||||||||||||||
Db    998 TATAAGTGTGTAGAGATTGTTCCTGGAAATTCATATTTGTGTGCGAAGACATATTACCCATGTGTATGCAA 1057

Qy   1629 CATCTACCTAGAGCTCTCAATGGGAGGTGCTCGAAGACATATTACCCATGTGTATGGCAA 1688
          ||||||||||| ||||| ||||||||| ||||||||||||||||||||||||||||||||
Db   1629 CATCTACCTAGAGCTCTCAATGGGAGGTGCTCGAAGACATATTACCCATGTGTATGCAA 1688
```

Db   1058 AAGTTGCCTAAAGCTTTGAATGGAAGAGATATTTGAAAAAATATTATCCCAGTATATGGTAG 1117

Qy   1689 GATGTTT 1695
               |||||||
Db   1118 GGAGCTT 1124

RESULT 10
US-10-425-115-159738
; Sequence 159738, Application US2004014272A1
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;   APPLICANT: La Rosa, Thomas J.
;   APPLICANT: Kovalic, David K.
;   APPLICANT: Zhou, Yihua
;   APPLICANT: Cao, Yongwei
;   TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;   TITLE OF INVENTION: Plants
;   FILE REFERENCE: 38-21(53222)B
;   CURRENT APPLICATION NUMBER: US/10/425,115
;   CURRENT FILING DATE: 2003-04-28
;   NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 159738
;   LENGTH: 5913
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_77253C.1
US-10-425-115-159738

Query Match   5.8%;   Score 155.6; DB 18; Length 5913;
Best Local Similarity   62.5%;   Pred. No. 7.1e-31;
Matches 305;   Conservative   0;   Mismatches 159;   Indels   24;   Gaps   3;

Qy   1232 GAGCTCATCATTTATATAAGCATGGTGATACCCAAAATTACTCCTTTTGAGCATGTTTATA 1291
               || | ||||||||| || | ||||||| |||||||||| |||||||||||||||||
Db   5426 GGGCCTCATAGAGATATCTAAACATAGGGGCTACTAAAGTATCTCCTTTTGAGCTTGTCTATG 5485

Qy   1292 TTTAGGAGGACGTTTTACCTGTTGAGGTAAATCGAACCTAATAAATCGGCTAAGCAAA 1351
               |||||||||||| ||||||||| |||||||||| ||||||||||||||||||||||
Db   5486 GGCAGGAGGCAGTGTTGCCTGTGTGAAATAAGTTTGAAATGCTATCAGGTTCGCCAGACAAA 5545

Qy   1352 ATAATTTATCACCTGTTGATTCTAACAATTGATGATGGACAATATTGATGAGGTGACTG 1411
               ||| || |||||||||| ||||||||||| |||||||||||||||||||||||||||
Db   5546 ATGATCTAACTGCTGCTACCGGATTATTATGATTCAATGATGATAATATTGATGAGGTGACTG 5605

Qy   1412 ACAAATGATTGAAGGCTTTAAAGGAGGATAA------------------TCT 1455
               |||| | || ||||| ||||||||||||||| |||
Db   5606 ACAAAAGGATGATAGCTTTGGGGACTATAGAAAAAGACAAAATCATGGTAGCCAGGGCCT 5665

Qy   1456 ACAATAAAAATGTAAAGAGAAAAGCATTCAAAGTGTGAGATCTGGTGTGGAAGACTATTT 1515
               |||||| ||| |||||||||||||||||||| |||||||||||| |||||||||||
Db   5666 ACAACAAAAAGGTCAAAGCAAATCTTTTCAAGTTGTGTTGGAAGACCATTC 5725

Qy   1516 TGCCTCTTCGGGGTAAAAGACAACAAGTTTAGTAAGTGGCCTCAAAATTGGGAGGGCCCA 1575
               |||||||||||| || ||||| | ||||||||||||| ||||| |||||||||||||||
Db   5726 TGCCTCTAAGGAGTAA-AGACCGGAAGTTCGGGAAATGTCACCAGCTGGGAAGGCCCT 5784

Qy   1576 TGCAAGATTGTTAAAGTAATTGTTTTGGA------TTGACGGAGGCATTTCAAGGTGAT 1628
               |||||||||||| |||||| ||||||||| |||||||||||||||||||||||
Db   5785 TACAAAGTAAAAACAGGTGATGTTTGGCACAGCCCTATTTGCTACAAACATTACAAGGCAAG 5844

Qy   1629 CATCTACCTAGAGCTTCCAATGGGAGGTGCTCGAAGACATATTACCCATGTGTATGGCAA 1688
               ||||||||||||||||||||||||| ||||||||||||| ||||||||||||||||||
Db   5845 GACTTGCCTAAGGCTTTGAATGGGCGTTTCCTCAAACAGTACCATCCTAGTATGTGGCAG 5904

Qy   1689 GATGTTTA 1696
               ||||| ||
Db   5905 GATGCCTA 5912

RESULT 11
US-10-437-963-25937
; Sequence 25937, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
;   APPLICANT: La Rosa, Thomas J.
;   APPLICANT: Kovalic, David K.
;   APPLICANT: Zhou, Yihua
;   APPLICANT: Cao, Yongwei
;   APPLICANT: Wu, Wei
;   APPLICANT: Boukharov, Andrey A.
;   APPLICANT: Barbazuk, Brad
;   APPLICANT: Li, Ping
;   TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;   TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;   FILE REFERENCE: 38-21(53221)B
;   CURRENT APPLICATION NUMBER: US/10/437,963
;   CURRENT FILING DATE: 2003-05-14
;   NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 25937
;   LENGTH: 4014
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_30775C.1
US-10-437-963-25937

Query Match   5.8%;   Score 155; DB 18; Length 4014;
Best Local Similarity   62.4%;   Pred. No. 8.2e-31;
Matches 305;   Conservative   0;   Mismatches 160;   Indels   24;   Gaps   3;

Qy   1232 GAGCTCATCATTTATATAAGCATGGTGATACCCAAAATTACTCCTTTTGAGCATGTTTATA 1291
               | ||| |||||||| || | ||||| |||||||||||||||||||||||||||||||
Db   3527 GGGGCCATAGGATATCTAAACATGGTGCCACTAAAGTCACTCCTTTTGAGTTGGTTTATG 3586

Qy   1292 TTTAGGAGGACGTTTTACCTGTTGAGGTAAATCGAACCTAATAAATCGGCTAAGCAAA 1351
               ||||||||||||| |||||||| || |||||| |||| ||||||||||||||||||
Db   3587 GTCAAGAAGCCGTTTTACCCGTGAGGTAAATCTTGGTTCTCGTTATATCAAGCAAG 3646

Qy   1352 ATAATTTATCACCTGTTGATTCTAACAATTGATGATGGACAATATTGATGAGGTGACTG 1411
               || | || ||||| || ||||||||| |||||||||||| ||||||||||||||||||
Db   3647 ATGATTTGTCAAGTGAAGATTACAAGACATTGATGGGAGACAACCTTGATGAAGTCATCG 3706

Qy   1412 ACAAATGATTGAAGGCTTTAAAGGAGGATAA------------------ATCT 1455
               ||||| || ||||| ||||||||||||||| |||
Db   3707 ACAAACGCTTGAAGGCATTGGAGGAGATAGAAAAGGAGAAGAGAGGGTGGCCAAGGCGT 3766

Qy   1456 ACAATAAAAATGTAAAGAGAAAGCATTCAAAGTGTGAGATCTGGTGTGGAAGACTATTT 1515
               |||||| ||| ||||||||||||||||||| | ||| || ||| ||||| |||||||
Db   3767 ACACCAAGAGGGTGAAAGCAAAATTGTTTCAAGTTCAGGTTGGGACTTGGGAAGACAATTT 3826

Qy   1516 TGCCTCTTCGGGGTAAAAGACAACAAGTTTAGTAAGTGGCCTCAAAATTGGGAGGGCCCA 1575
               ||| | |||| || ||| ||| || ||||||||||| |||| | |||||||||||||||
Db   3827 TGCCT-TTGGGTACTCGATCCAAGGAGTTCCGTAACGTGGTGGTCCTCCTAGTTGGGAAGGTCCT 3885

Qy   1576 TGCAAGATTGTTAAAGTAATTGTTTTGGA------TTGACGGAGGCATTTCAAGGTGAT 1628
               ||||||||||| ||||| |||||||||| |||||||||||||||||||||||
Db   3886 TATCGAGTGTGCGGCATTGTTCGACGTGAACGCATACTTTTTGGAGACACTTCAAGGGGAG 3945

Qy   1629 CATCTACCTAGAGCTTCCAATGGGAGGTGCTCGAAGACATATTACCCATGTGTATGGCAA 1688
               |||||||||||||||| ||||||| || ||| || |||| ||||||||||||||||||
Db   3946 CGTTTTCAGCGGAGCAATCAACCACGGAAATATACTTGAAGAAATACTTCCCGAGTGTTTGTCAA 4005

Qy   1689 GATGTTTAG 1697
               ||||| |||
Db   4006 GACGCTTAG 4014

RESULT 12
US-10-437-963-26057/c
; Sequence 26057, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
;   APPLICANT: La Rosa, Thomas J.
;   APPLICANT: Kovalic, David K.

```
;     APPLICANT:  Zhou, Yihua
;     APPLICANT:  Cao, Yongwei
;     APPLICANT:  Wu, Wei
;     APPLICANT:  Boukharov, Andrey A.
;     APPLICANT:  Barbazuk, Brad
;     APPLICANT:  Li, Ping
;     TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;     TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;     FILE REFERENCE: 38-21(53221)B
;     CURRENT APPLICATION NUMBER: US/10/437,963
;     CURRENT FILING DATE: 2003-05-14
;     NUMBER OF SEQ ID NOS: 204966
;     SEQ ID NO 26057
;     LENGTH: 7495
;     TYPE: DNA
;     ORGANISM: Oryza sativa
;     FEATURE:
;     OTHER INFORMATION: Clone ID: PAT_MRT4530_30884C.1
US-10-437-963-26057

  Query Match       5.8%;
  Best Local Similarity    63.7%;    Score 155;  DB 18;  Length 7495;
                                     Pred. No. 1.2e-30;
  Matches  313;  Conservative   0;  Mismatches  150;  Indels  28;  Gaps   4;

Qy   1232 GAGCTCATCATTTATATAAGACATGTGATACCAAAATTACTCCTTTGAGCATGTTTATA 1291
          ||||| |||| ||||||| |||||||   ||||||||||||||||||||||| |||||
Db    488 GGGCACATAGGATATCTAAACATGGTGCTACTAAAGTCACTCCTTTGAGTTGGTTTATG 429

Qy   1292 TTTAGGAGACGTTTTACCTGTTGAGGTAAATCTGAACGCTAATAAATCGGCTAAGCAAA 1351
          ||||                                              ||||||||
Db    428 GTCAAGAAGCCGTTTACCGGTGGAGGTAAATCTTTGTCTCTTCGTTATATCAAGCAAG 369

Qy   1352 ATAATTTATCACCTGTGATTCTAACAATTTGATGATGGACAATATTGATGAGGTGACTG 1411
          ||||||||                              |||||||||||||||||||
Db    368 ATGATTTGTCAAGTGAAGATTACAAGACATTGGGACAACCTTGATGAAGTCATCG 309

Qy   1412 ACAAATGATTGAAAGCTTAAAGGAGGATTGAGAGAGATAA----------ATCT 1455
          |||| |||||||| || ||||                              ||||
Db    308 ACAAGCGCTTGAAGGCATTCGAAGCATTGAGAAGAAGAGAGAGGTGGCCAAGGCGT 249

Qy   1456 ACAATAAAAATGTAAAGAAGAAAGCATTCAAAGTGTGAGATCTCGTGTGGAAGACTATTT 1515
          |||| |||  |||||||| |||| |||||||||| |||||||  ||  ||||||
Db    248 CCAACAAAAGGGTGAAAGGAAAGTTGTTTCAAGTTGGAGATTGGTTTGCTTGGAACAATTT 189

Qy   1516 TGCCTCTTGGGGGGTAAAAGCAACAAGTTTAGTAAGTGGCCTCAAAATTGGGAGGGCCC- 1574
          |||||  |||||  |       |||||||||||||||||  |  ||||||||||||||
Db    188 TGCCT-TTGGGTACTCGATCCAAGGAGTTCGGTAAGTGGTTTCCTAGTTGGGAGGGCCCT 130

Qy   1575 ------ATGCAAGATTGTTAAAGTAATTGTTTTGGATTGACGGAGGCATTCAAGGTG 1626
                  | ||  ||||||| ||||| |    |||||||||           |||||||
Db    129 TATCGAAATAGCGGCATTGTTCGAGAAATGCATATTTTTA--GAGACACTTCAAGGGG 72

Qy   1627 ATCATCTACCTAGAGCTCTCAATGGGAGGTGCTCGAAGACATATTACCCATGTGTATGC 1686
          |  || |||| |||||| ||||  |||   |||| |||||||| |||| ||||| ||||
Db     71 AGCGTTTCAGCGAGCAATCAACGGGAAATACTTGAAGAAATACTTCCCGAGCGTTTGGC 12

Qy   1687 AAGATGTTTAG 1697
          |||| |||||
Db     11 AAGAGCGTTAG 1


RESULT 13
US-10-437-963-85935
; Sequence 85935, Application US10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;     APPLICANT:  La Rosa, Thomas J.
;     APPLICANT:  Kovalic, David K.
;     APPLICANT:  Zhou, Yihua
;     APPLICANT:  Cao, Yongwei
;     APPLICANT:  Wu, Wei
;     APPLICANT:  Boukharov, Andrey A.
;     APPLICANT:  Barbazuk, Brad
```

```
;     APPLICANT:  Li, Ping
;     TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;     TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;     FILE REFERENCE: 38-21(53221)B
;     CURRENT APPLICATION NUMBER: US/10/437,963
;     CURRENT FILING DATE: 2003-05-14
;     NUMBER OF SEQ ID NOS: 204966
;     SEQ ID NO 85935
;     LENGTH: 5928
;     TYPE: DNA
;     ORGANISM: Oryza sativa
;     FEATURE:
;     OTHER INFORMATION: Clone ID: PAT_MRT4530_85023C.1
US-10-437-963-85935

  Query Match       5.7%;
  Best Local Similarity    62.8%;    Score 154.6;  DB 18;  Length 5928;
                                     Pred. No. 1.3e-30;
  Matches  301;  Conservative   0;  Mismatches  154;  Indels  24;  Gaps   3;

Qy   1232 GAGCTCATCATTTATATAAGACATGTGATACCAAAATTACTCCTTTGAGCATGTTTATA 1291
          |||||||||||  |||| || |||| ||| |||  ||||| ||||||| |||  |||
Db   5450 GGGCTCATCATCGGATATCTAAGCATGGTGCGACTCAAGTAACACCGTATGAGCTAGTCTATG 5509

Qy   1292 TTTAGGAGACGTTTTACCTGTTGAGGTAAATCTGAACGCTAATAAATCGGCTAAGCAAA 1351
          |  || | ||||| |||| ||||||||||     || || || ||||  || |||||||
Db   5510 GGCAGAAGGCCGTGTTACTTGTGGAGGTAATTTACAAGCTCTTAGAATAGCCAGACAAA 5569

Qy   1352 ATAATTTATCACCTGTGATTCTAACAATTTGATGATGGACAATATTGATGAGGTGACTG 1411
          ||  ||| || ||||| ||||| | ||||| | |||||||| ||  ||||| || ||
Db   5570 ATGGTTTGTCGGCTGAGATTTATAGGAATTTAATGATGGACAGGATAGATGACGTCCCAG 5629

Qy   1412 ACAAATGATTGAAAGCTTAAAGAGAGATTGAGAAGGATAAAT------------CT 1455
          | |||  |||||| ||||||| |  || ||||||||  ||||             ||
Db   5630 AGGAAAGATTGAAGGCCTTGAGACATATTGAGAAGAAAAATTAAGAGTAGCTAGGGCTT 5689

Qy   1456 ACAATAAAAATGTAAAGAAGAAAGCATTCAAAGTGTGAGATCTCGTGTGGAAGACTATTT 1515
          | |||||  | ||| || ||| ||||||| ||   |||||    |||| ||| ||||||
Db   5690 ATAATAAGAGGTGAGAGAAAAATCATTCAGATAGGTGAGTTGGTATGGAAAACGATTT 5749

Qy   1516 TGCCTCTTGGGGGGTAAAAGCAACAAGTTTAGTAAGTGGCCTCAAAATTGGGAGGGCCCA 1575
          ||||  |||| |  || |   |||||| |  | |||||| |  | || ||| || ||
Db   5750 TGCCCATTGGAACTAGGAG-TAGCAAATTTGGCAAGTGGTCACCGAGCTGGGAAGTCCT 5808

Qy   1576 TGCAAGATTGTTAAAGTAATTGTTTTGGATTGA------CGGAGGCATTCAAGGTGAT 1628
          | |||||||||| |||  |||||    ||||||       |  |  |||||| ||||
Db   5809 TATAAAGTTGTAGAGATTGTTCCTCGGAAATTCATATTTGTGCAGTCTTTGCAGGGGGAT 5868

Qy   1629 CATCTACCTAGAGCTCTCAATGCGGAGGTGCTCGAAGACATATTACCCATGTGTATGCA 1687
          || ||  ||| ||||| |||||  ||||    | |||| |||||| |||| ||| |||
Db   5869 AAGTTGCCTAAAGCTTTGAATGAAGGAAGATATTTGAAAAATATTATCCCAGTATATGGTA 5927


RESULT 14
US-10-425-115-54555
; Sequence 54555, Application US2004021427A1
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;     APPLICANT:  La Rosa, Thomas J.
;     APPLICANT:  Kovalic, David K.
;     APPLICANT:  Zhou, Yihua
;     APPLICANT:  Cao, Yongwei
;     TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;     TITLE OF INVENTION: PlantS
;     FILE REFERENCE: 38-21(53222)B
;     CURRENT APPLICATION NUMBER: US/10/425,115
;     CURRENT FILING DATE: 2003-04-28
;     NUMBER OF SEQ ID NOS: 369926
;     SEQ ID NO 54555
;     LENGTH: 5460
;     TYPE: DNA
;     ORGANISM: Zea mays
;     FEATURE:
;     OTHER INFORMATION: Clone ID: MRT4577_149755C.1
```

US-10-425-115-54555

Query Match   5.7%;   Score 154;   DB 18;   Length 5460;
Best Local Similarity   62.3%;   Pred. No. 1.9e-30;
Matches 304;   Conservative   0;   Mismatches 160;   Indels 24;   Gaps   3;

```
Qy   1232 GAGCTCATCATTTATATAAGCATGGTGATACCAAAATTACTCCTTTTGAGCATGTTTATA 1291
          | ||||| ||| ||| || || |||| ||  ||| |||||| ||||| ||||
Db   4973 GGGCTCATAGAATATCTAAACATAGGGCTACTAAAGTATCTCCTTTTGAGCTTGTCTATG 5032

Qy   1292 TTTAGGAGGACGTTTACCTGTTGAGGTAAATCTGAACGCTAATAAATCGGCTAAGCAAA 1351
          |||| || ||| ||| |||||  ||| |||| |||| ||  ||  |||||
Db   5033 GGCAGGAAGCAGTGTTACCTGTTGCCAGACAAA 5092

Qy   1352 ATAATTATCACCTGTTGATTCTAACAATTTGATGATGACAATATTGATGAGGTGACTG 1411
          ||||| ||| |||||| ||| |  |||| ||||| | ||| |||||||||||||||||
Db   5093 ATGATCTAACTGTTACTAATTATTATAATTCAATGATGATAATATTGATGAGGTGACCG 5152

Qy   1412 ACAAATGATTGAAGGCTTTAAAGGAGATTGAGAAGGATAA-----------ATCT 1455
          |||| ||| || || ||| |||| |||| || |||| ||
Db   5153 ACAAGAGGATGATAGCTTTGGGAGCAATAGAAAAGGACATCATGGTGAGCAGGGCCT 5212

Qy   1456 ACAATAAAAATGTAAAGAAGAAAGCATTCAAAGTGAGATCTGGTGTGGAAGACTATTT 1515
          ||||| ||| || ||| ||| ||| ||||| ||||||||||||||||||||||||||
Db   5213 ACAACAAGAGGTCAAAGCAAAATCATTCCAGTAGGAGACTTGGTGTGGAAGACCATTC 5272

Qy   1516 TGCCTCTTGGGGTAAAAGACAACAAGTTTAGTAGTGGCCTCAAAATTGGGAGGGCCCA 1575
          |||||  ||| |||||| |||||| ||||| |||  | |  ||||||||||||||
Db   5273 TACCTTCTAAGGAAT-AAAGATCGAAAGTTCGGGAAATGGTCGCCAAGCTGGGAGGGTCCT 5331

Qy   1576 TGCAAGATTGTTAAAGTAA-------TTGTTTTGGATTGACGGAGGCATTTCAAGTGAT 1628
          ||| | ||| |  |  ||         ||| |  | |  |  | |  ||||| | |
Db   5332 TATAAAGTGAAACAGGTGATGTGTCGGTAATGCTTATTTACTACAACATTACAAGGCAAG 5391

Qy   1629 CATCTTACCTAGAGCTCTCAATGGAGGTGCTCGAAGACATATTACCCATGTGTATGGCAA 1688
          | | | |||  | |  |  | || |||| |   ||||||||||||||||||||||||||
Db   5392 GATTTACCTAAGGCTTTGAATGGGCGTTTCCTCAAACAGTACCATCCTAGTATGTGGCAA 5451

Qy   1689 GATGTTTA 1696
          ||||
Db   5452 GATGCCTA 5459
```

RESULT 15
US-10-425-115-159747
US-10-425-115-159747

Query Match   5.7%;   Score 154;   DB 18;   Length 7170;
Best Local Similarity   62.3%;   Pred. No. 2.2e-30;
Matches 304;   Conservative   0;   Mismatches 160;   Indels 24;   Gaps   3;

```
Qy   1232 GAGCTCATCATTTATATAAGCATGGTGATACCAAAATTACTCCTTTTGAGCATGTTTATA 1291
          | ||||| ||| ||| || || |||| ||  ||| |||||| ||||| ||||
Db   6683 GGGCTCATAGAATAATCTAAACATAGTGCTACTAAAGTATCTCCTTTTGAGCTTGTCTATG 6742
```

```
Qy   1292 TTTAGGAGGACGTTTACCTGTTGAGGTAAATCTGAACGCTAATAAATCGGCTAAGCAAA 1351
          |||| || ||| ||| |||||  ||| |||| |||| ||  ||  |||||
Db   6743 GGCAGGAAGCAGTGTTACCTGTTGCCAGACAAA 6802

Qy   1352 ATAATTATCACCTGTTGATTCTAACAATTTGATGATGACAATATTGATGAGGTGACTG 1411
          ||||| ||| |||||| ||| |  |||| ||||| | ||| |||||||||||||||||
Db   6803 ATGATCTAACTGTTACTAATTATTATAATTCAATGATGATAATATTGATGAGGTGACCG 6862

Qy   1412 ACAAATGATTGAAGGCTTTAAAGGAGATTGAGAAGGATAA-----------ATCT 1455
          |||| ||| || || ||| |||| |||| || |||| ||
Db   6863 ACAAGAGGATGATAGCTTTGGGAGCAATAGAAAAGGACATCATGGTGAGCAGGGCCT 6922

Qy   1456 ACAATAAAAATGTAAAGAAGAAAGCATTCAAAGTGAGATCTGGTGTGGAAGACTATTT 1515
          ||||| ||| || ||| ||| ||| ||||| ||||||||||||||||||||||||||
Db   6923 ACAACAAGAGGTCAAAGCAAAATCATTCCAGTAGGAGACTTGGTGTGGAAGACCATTC 6982

Qy   1516 TGCCTCTTGGGGTAAAAGACAACAAGTTTAGTAGTGGCCTCAAAATTGGGAGGGCCCA 1575
          |||||  ||| |||||| |||||| ||||| |||  | |  ||||||||||||||
Db   6983 TACCTTCTAAGGAAT-AAAGATCGAAAGTTCGGGAAATGGTCGCCAAGCTGGGAGGGTCCT 7041

Qy   1576 TGCAAGATTGTTAAAGTAA-------TTGTTTTGGATTGACGGAGGCATTTCAAGTGAT 1628
          ||| | ||| |  |  ||         ||| |  | |  |  | |  ||||| | |
Db   7042 TATAAAGTGAAACAGGTGATGTGTCGGTAATGCTTATTTACTACAACATTACAAGGCAAG 7101

Qy   1629 CATCTTACCTAGAGCTCTCAATGGAGGTGCTCGAAGACATATTACCCATGTGTATGGCAA 1688
          | | | |||  | |  |  | || |||| |   ||||||||||||||||||||||||||
Db   7102 GATTTACCTAAGGCTTTGAATGGGCGTTTCCTCAAACAGTACCATCCTAGTATGTGGCAA 7161

Qy   1689 GATGTTTA 1696
          ||||
Db   7162 GATGCCTA 7169
```

Search completed: May 9, 2005, 05:14:09
Job time : 1604 secs

THIS PAGE LEFT BLANK

```
                Copyright  GenCore version 5.1.6
                Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:   May 8, 2005, 21:19:18 ; Search time 11588 Seconds
                                  (without alignments)
                                  11248.234 Million cell updates/sec

Title:          US-10-751-550-1
Perfect score:  2690
Sequence:       1 tctagataatacgactcact.........ccataactagtaccaccatg 2690

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       4708233 seqs, 2422760795 residues

Total number of hits satisfying chosen parameters:   9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  GenEmbl:*
             1: gb_ba:*
             2: gb_htg:*
             3: gb_in:*
             4: gb_om:*
             5: gb_ov:*
             6: gb_pat:*
             7: gb_ph:*
             8: gb_pl:*
             9: gb_pr:*
            10: gb_ro:*
            11: gb_sts:*
            12: gb_sy:*
            13: gb_un:*
            14: gb_vi:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	C	Score	Query Match	Length	DB	ID	Description
1		1030.2	38.3	1963	8	AJ626722	AJ626722 Saccharum
2		192.4	7.2	107491	8	AP004613	AP004613 Oryza sat
3		192.4	7.2	113514	8	AP004767	AP004767 Oryza sat
4		188.6	7.0	144596	8	AP004574	AP004574 Oryza sat
5		188.6	7.0	165766	8	AP004706	AP004706 Oryza sat
6		182.8	6.8	123269	2	OSJN00028	AL606596 Oryza sat
7		182.8	6.8	176627	8	AP006062	AP006062 Oryza sat
8	c	181.4	6.7	139201	8	AP002970	AP002970 Oryza sat
9	c	181.4	6.7	150022	8	AP003213	AP003213 Oryza sat
10		175.8	6.5	136858	8	AC134345	AC134345 Oryza sat
11		175.8	6.5	144294	8	AP004663	AP004663 Oryza sat
12		175.8	6.5	149435	8	AC136521	AC136521 Oryza sat
13		175.8	6.5	154416	8	AP005389	AP005389 Oryza sat
14		173.6	6.5	127117	8	OSJN00040	AL606603 Oryza sat
15		172	6.4	150137	8	AC135226	AC135226 Oryza sat
16	c	171.2	6.4	198867	8	AP005325	AP005325 Oryza sat
17	c	170.8	6.3	130521	8	AL606589	AL606589 Oryza sat
18	c	170.8	6.3	145766	8	OSJN00287	AL731644 Oryza sat
19	c	170.4	6.3	102515	8	H0806H05	AL442113 Oryza sat
20		170.2	6.3	121591	8	OSJN00120	AL606989 Oryza sat
21		169.6	6.3	137779	8	AP005843	AP005843 Oryza sat
22		169.6	6.3	149581	8	AP005484	AP005484 Oryza sat
23		169.6	6.3	177529	8	AP005305	AP005305 Oryza sat
24	c	169.2	6.3	128574	8	OSJN00199	AL663002 Oryza sat
25		169.2	6.3	152702	8	AP005162	AP005162 Oryza sat
26	c	169.2	6.3	160462	8	OSJN00172	AL662971 Oryza sat
27		169.2	6.3	184235	8	AP005166	AP005166 Oryza sat
28		169.2	6.3	187972	8	AY360390	AY360390 Oryza sat
29	c	166.8	6.2	138715	8	OSJN00190	AL662989 Oryza sat
30		166.2	6.2	137297	8	AC090714	AC090714 Oryza sat
31		164.8	6.1	171790	8	AP006460	AP006460 Oryza sat
32		162.8	6.1	148373	8	AC134348	AC134348 Oryza sat
33	c	162.6	6.0	4239	8	AK069351	AK069351 Oryza sat
34		162.6	6.0	117057	8	AP004180	AP004180 Oryza sat
35		162.6	6.0	148060	8	AP003334	AP003334 Oryza sat
36	c	161.8	6.0	120640	8	OSJN00113	AL606991 Oryza sat
37	c	161.8	6.0	131457	8	AC144718	AC144718 Zea mays
38	c	161.6	6.0	136214	8	AC134517	AC134517 Oryza sat
39		161.6	6.0	169500	8	AC115686	AC115686 Oryza sat
40	c	161.6	6.0	300029	8	AE017077	AE017077 Oryza sat
41		161	6.0	176150	8	AC135864	AC135864 Oryza sat
42		160.4	2	146528	8	AP005507	AP005507 Oryza sat
43	c	160.4	6.0	158749	8	AY360394	AY360394 Oryza sat
44		160.4	6.0	160541	8	AP004458	AP004458 Oryza sat
45		160.4	6.0	188058	8	AP005819	AP005819 Oryza sat

ALIGNMENTS

```
RESULT 1
AJ626722
LOCUS       AJ626722               1963 bp    DNA     linear   PLN 06-FEB-2004
DEFINITION  Saccharum officinarum partial dpb gene for putative dirigent
            protein.
ACCESSION   AJ626722
VERSION     AJ626722.1  GI:42454401
KEYWORDS    dirigent protein; dpb gene.
SOURCE      Saccharum officinarum
  ORGANISM  Saccharum officinarum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum
            complex.
REFERENCE   1
  AUTHORS   Abraha,T.G., Rogbeer,O., Groenewald,S., Groenewald,J.-H. and
            Botha,F.C.
  TITLE     Characterisation of culm-specific promoter elements in sugarcane
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1963)
  AUTHORS   Abraha,T.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-2004) Abraha T.G., Botany, Institute for Plant
            Biotechnology, Private Bag X1, Matieland, 7602, SOUTH AFRICA
FEATURES             Location/Qualifiers
     source          1..1963
                     /organism="Saccharum officinarum"
                     /mol_type="genomic DNA"
                     /variety="N19"
                     /isolation_source="Welgevallen experimental farm"
                     /db_xref="taxon:4547"
                     /clone="pGUS-DPB"
                     /tissue_type="culm"
                     /country="South Africa:Stellenbosch"
     gene            1..1715
                     /gene="dpb"
     promoter        <1..1151
                     /gene="dpb"
                     /function="guiding protein"
     CDS             1152..1715
                     /gene="dpb"
                     /codon_start=1
```

```
          /product="putative dirigent protein"
          /protein_id="CAP25234.1"
          /db_xref="GI:42454402"
          /translation="MASLRSVLAVSLALALFAVAPASFALDERELHSLYLNQTYSGN
          GLNQAVVVEPGLPGEFGNNAIQDWPVTNGEGSDATVVGRAQGIQFKPSERNDQAWYT
          LTIVFERTSLKGSTLQMGYIFQDGQWSIFGGTGQLTMARGVVNHKVVRQTNGRRMYK
          INIYAFYTPLGASSSCRINPRRLDFDA"

ORIGIN

Query Match    38.3%;      Score 1030.2;   DB 8;     Length 1963;
Best Local Similarity   94.9%;   Pred. No. 4.3e-246;
Matches 1136;   Conservative   0;   Mismatches   13;   Indels   48;   Gaps   5;

Qy   1490 GTGAGATCGGTGGTGGAAGACTATTTGCCTCTTGGGGGTAAAAGACAACAAGTTTAGTA 1549
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GTAAGATCTGGTGTGGAAGACTATTTGCCTCTTGGGGGTAAAAGACAACAAGTTTAGTA 60

Qy   1550 AGTGGCCTCAAAATTGGGAGGGGCCCATGCAAGATTGTAAAGTAATTGTTTTGGATTGAC 1609
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AGTGGCCTCAAAATTGGGAGGGGCCCATGCAAGATTGTAAAGTAATTGTTTTGGATTGAC 120

Qy   1610 GGAGGCATTTCAAGTGATCATCTACCTAGAGCTCTCAATGGGAGGTGCTCGAAGACATA 1669
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGAGGCATTTCAAGTGATCATCTACCTAGAGCTCTCAATGGGAGGTGCTCGAAGACATA 180

Qy   1670 TTACCCATGTGTATGGCAAGATGTTAGCTAGTAACTGACTGATAGTGTAAACGATCTCC 1729
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TTACCAAGTGTATGGCAAGATGTTAGCTAGTAACTGACTGATAGTGTAAACGATCATCTCC 240

Qy   1730 AATGGGGCAAGACATATTACCTAGGCCAGGCTGGTTTTCGAAGTTCGAGTAGGATATA 1789
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AATGGGGCAAGACATATTACCTAGGCCAGGCTAGTTTTCGAAGTTCAAGTAGGATATA 300

Qy   1790 GAGATTCTCGTGCGCAGTTGTAAACGATCTCCAATGGGCAAGACATCCTACCCTATATAT 1849
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GAGATTCTCGTGCGCAGTTGTAAACGATCTCCAATGGGCAAGACATCCTACCCTATATAT 360

Qy   1850 AGTGAAGGGGCAGTAGCTGATTGAGAATCAAGCACAATATAATTATTAATT 1909
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AGTGAAGGGGCAGTAGCTGATTGAGAATCAAGCACAATATAATTATTAATT 420

Qy   1910 TTTTATTCAAACCCAATTTTTTCCTTTTCCAACCCTAATTATAGTTTTCCTTTTTTGCCTC 1969
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TTTTATTCAAACCCAA--ATTTTTCCTTTTCCAACCCTAATTATAGTTTCCCTTTTTTGCCTC 479

Qy   1970 TAGGACAAATTGACGTGTTCCGGGTATCCTGCTGAATTAAGAACAACCCCTAGGTGCACCT 2029
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    480 TAGGACAAATTGACGTGTTCCAGTATCCTGCTGTAATTCAAGAACAACACCCTAGGTGCACCT 539

Qy   2030 GTCCCGATAGAGTCCCACCTGGGTAGGCATTCATAGGG-ATTCGTGTATTCCTGC-AAA 2087
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    540 GTCCCGATAGAGTCCCACCTGGGTAGGCATTCATAGGGAATTCGGTATTCCTGCAAAA 599

Qy   2088 AAAGCGATTAAGCTGGCTTCTAAAACTGGCTAGGCCGGATTCGTGCCTTCACTACCAG 2147
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    600 AAAGCGATTAAGCTGGCTCTAAAACTGGCTAGGCCGGATTCGTGCCTTCACTACCAG 659

Qy   2148 GTGATTTCATGTGATCCGTGCATTCTAGACACTTGCTATGTAACCCAAACTTAAGTCGA 2207
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    660 GTGATTTCATGTGATCCGTGCATTCTAGCACTTGCTATGTAACCCAAACTTAAGTCGA 719

Qy   2208 CAACTATAAATAATGCTACTTGCAGGATGTTATCACGACACAACAACTCCTAATCTACGGAAGC 2267
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    720 CAACTATAAATAATGCTACTTGCAGGATGTTATCACGACACAACAACTCCTAATCTAC-GAAGC 778

Qy   2268 CTAAGTTTAGTTTTGCTTCGGAGACAAGCAATTGTGGCCAGTCACTAGCTAGCTACGTTCAGAG 2327
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    779 CTAAGTTTAGTTTTTGTTCGGAGACAGCAATTGTGGCCAGTCACTATAGCTACGTTCAGAG 838

Qy   2328 GGTAGTGGGAGCAGTTGCGTCGTGTGGATTGA------------------ 2387
               ||||||||||||||||||||||||||
Db    839 GGTAGTGGGAGCAGTTGCGTCGTGTTGCGTCGTTGGATTGA 869
```

```
Qy   2388 ATTCACATGGACAGTAAATGGTACAGTAACTTCGCAAACAATAAAATCGTCACAATT 2447
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    870 --------------AAATGGTACAGTAACTTCGCAAACAATAAAATCGTCACAATT 914

Qy   2448 TATTAGTGCACTCCTCGACGTAAATGCTTCTCACGTCAGAGGATTTGATTCCGAGGGCCG 2507
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    915 TATTAGTGCACTCCTCGACGTAAATGCTTCTACGTCAGAGGATTTGATTCCGAGGGCTG 974

Qy   2508 CTGCACCACCAATGACGGTCTTTACCCATCCATCATGACCATTGTTCACATCCATG 2567
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    975 CTGCACCCATCACTAATGACGGTCTTTACCCATCCATCATGACCATTGTTCACATCCATG 1034

Qy   2568 CTATCACGTGTCGTCCTGTCCATGCACTGCACGCCCCTCTATAAAATACTGGCATCCCTCCCCC 2627
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1035 CTATCACGTGTCGTCCTGTCCAAGCACTGCACGCCCCTCTATAAAATACTGGCATCCCTCCCCC 1094

Qy   2628 GTTCACAGATCACCACAACCAAGCAAGAGAAATAAACGGTAGCTGCCATAACTAGTACA 2684
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1095 GTTCACAGATCACCACAACCAAGCAAGAAGATAAACGGTAGCTGCCATAACTAGTACA 1151
```

```
RESULT 2
AP004613
LOCUS       AP004613              107491 bp    DNA     linear   PLN 31-AUG-2004
DEFINITION  Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
            BAC clone:OJ1123_G09.
ACCESSION   AP004613 BA000010
VERSION     AP004613.3  GI:21128156
KEYWORDS    .
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC
            clone:OJ1123_G09
  JOURNAL   Published Only in Database (2002)
REFERENCE   2  (bases 1 to 107491)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JAN-2002) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki, 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     On Jun 6, 2002 this sequence version replaced gi:19773520.
            Genes were predicted from the integrated results of the following:
            GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor
            (October 1998 version). The genomic sequence was searched against
            NCBI NonRedundant Protein database, nr
            (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
            RGP. Protein homologies of the coding regions were searched against
            NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent
            the identified cDNA sequences using BLASTN 2.0 with the
            corresponding DDBJ accession no. and RGP clone ID.
            A gene with identity or significant homology to a protein.is
            classified based on the protein name to indicate the homology level
            such as same name, 'putative' and '.like protein'. A gene without
            significant homology to any protein but with EST homology (covering
            almost the entire length of partial sequence) is classified as an
            'unknown' protein. A gene predicted with a gene prediction program
            is classified as a 'hypothetical' protein.
            The orientation of the sequence is from -2lM13 to M13rev of the BAC
            clone. This sequence of OJ1123_G09 clone has an overlap with
            OJ1460_H08 (DDBJ: AP004767) at the position 86,928 to 107,491 of 3'
            end. The nucleotide sequence of this BAC clone was generated by
            combining Monsanto and RGP-Japan sequencing data. Detailed
            information on overlap and assembly quality together with
            annotation of this entry is available at
            http://rgp.dna.affrc.go.jp/GenomeSeq.html.
FEATURES             Location/Qualifiers
```

```
source          1..107491
                /organism="Oryza sativa (japonica cultivar-group)"
                /mol_type="genomic DNA"
                /cultivar="Nipponbare"
                /db_xref="taxon:39947"
                /chromosome="1"
                /clone="OJ1123_G09"
gene            complement(3..1367)
                /gene="OJ1123_G09.1"
misc_feature    complement(3..1367)
                /gene="OJ1123_G09.1"
                /note="probably inactive due to stop codon(s) in CDS
                pseudogene, gag-pol precursor"
                /pseudo
gene            complement(1601..2944)
                /gene="OJ1123_G09.2"
CDS             complement(1601..2944)
                /gene="OJ1123_G09.2"
                /note="hypothetical protein
                similar to Oryza sativa chromosome 10,OSJNBa0011A24.17"
                /codon_start=1
                /protein_id="BAC00734.1"
                /db_xref="GI:21328157"
                /translation="MSYYVDPRTSVPQYPLYSVYGYPPSTVACQVGGLGAQGSTMTS
                SMSKNPAAEAEIGTSTTSEPEKDPSAAKPCLSNRNHEPMRITSEVTRSWCPIHKTRKH
                TLQAWMVFLNVHAEIRACKECGIQHTSPTRDVHGPIHKTKNHDLSSCKVFLSAMKTPP
                PKVQQSRIPIRDEDKEQGATPILDRFVGVIDIDHEPSVLHLLEDYGSSTTSTPREVL
                AIDDVGTSVRTNAEAENQVTAPTQHIRAINAILRETPYDPILNDDQARWIERLQESVT
                NLSNVFEEAATAAHPEQPTGDANGENPERRESPHRATPPPRGTCDLLDHLNGRRRAR
                HTRDNENRSRHVSSRAIQAKIEEAIQAKTETVITTITTMIANGGAGRYWSRTSP"
gene            3436..4231
                /gene="OJ1123_G09.3"
misc_feature    3436..4231
                /gene="OJ1123_G09.3"
                /note="probably inactive due to stop codon(s) in CDS
                pseudogene, transposable element"
                /pseudo
gene            join(4645..5386,6323..6438)
                /gene="OJ1123_G09.4"
CDS             join(4645..5386,6323..6438)
                /gene="OJ1123_G09.4"
                /note="hypothetical protein
                similar to Oryza sativa chromosome 1, P0043B10.18"
                /codon_start=1
                /protein_id="BAC00735.1"
                /db_xref="GI:21328158"
                /translation="MHKENQVMESSLETKLIDMPQPQETNGDGERAPEADISMATTLM
                PNGVGERXATAAPALPAARPPEMTDKVMASTANLAQLLPTGTALAYQALSTSFTNHG
                QCYRSNRWLTAGLVAVLTASSIFFSLTDSVVGRGGKLYYGMATPRGFNVFNLSREEBE
                AQELSTRKLRELRVRPLDIVHAFFTAVVFLTVAFSDVGLTKCFFPDAGNDTKELLKNL
                PLGMAFMSTFVFLLFPFKRKGIGYTDTTPRIIDGINKTNFFDTAPTGKNFCSGVGEKR
                ARLSTTHCG"
gene            join(7843..8274,8373..8505,8786..8946)
                /gene="OJ1123_G09.5"
CDS             join(7843..8274,8373..8505,8786..8946)
                /gene="OJ1123_G09.5"
                /note="hypothetical protein"
                /codon_start=1
                /protein_id="BAC00736.1"
                /db_xref="GI:21328159"
                /translation="MASGGGWMRSPATATGGAGGGGQRGQRRRPKVTAVHARRTAVH
                TGERERGRRGEPHRVWRPERKAKAAVMIGDAKGRTGGDGTIGERGEGLGLGRKPAAT
                KNGQPEMERMEGKLWRSARKFGAHLTSTSVHSAPGSENFFYKCSSTRTLDKEGSSCTS
                IRVVINNLDVVSRLEVTFVCCVSHGFVRGEVLQRYSVLLTTKFGNFSIGDGDKTRAET
                KMEIVPESERGSSKVRIGCDWDR"
gene            join(12918..13037,13203..13586,14357..14887,15596:..16339)
                /gene="OJ1123_G09.6"
CDS             join(12918..13037,13203..13586,14357..14887,15596..16339)
                /gene="OJ1123_G09.6"
                /codon_start=1
                /product="putative polyprotein"
                /protein_id="BAC00737.1"
                /db_xref="GI:21328160"
                /translation="MPPKKVAKENVAARKESEAGVEMAAEEGAEPSALGAEADVAASP
                QQAIASTAHSTPEPTSKRKPKPTWGPCVRIWCDSRTCFVCKSSSRHTKWPGKPRSPR
                PPSSIRQATSHPRSTRKCRFKPFRHKSCKHQRTSPDSLRPPQYIRPLMYSHKFGRSL
                PKYAPRHLPKTEQHLLGIRQPRDSIKEYMCRFSQARCQVQDITEASIINAASAGLLE
                GELTRKIANKEPQTVEHLLRIIDEYARGEEDSKRRQAIQAEYDKASVIAAQAQVQAQV
                AEPTPLAVCQSQPANQGQPQRSGQAPMTWRKFRTDSAGKAVMAVEEVQALRKEFDAQQ
                ATSHQQPVRKKLKMPGPAGVIVKGLQPSATSKGDLAAINRAVHDVEAEPHDRAKHAP
                KPAPHGRVIKEQVNDADSTRFVSLGGMGEREVENILEVLKKVNIDIFTWGPDEVGGVS
                TDLIMHHLAVKPDARLRKQKLRKMSADRQEAAKAEVQKLKAGVIQEIDHPEWLANPV
                LVRKSNGKWRMCVDFTDLNKAYPKDDFPLPWIDQLVDSTAGCELMSFLDAYSGYHQIH
                MNPADIPKTAFITPSFTPCHSGCLSA"
gene            join(21504..21802,21856..22234)
                /gene="OJ1123_G09.7"
CDS             join(21504..21802,21856..22234)
                /gene="OJ1123_G09.7"
                /note="hypothetical protein"
                /codon_start=1
                /protein_id="BAC00738.1"
                /db_xref="GI:21328161"
                /translation="MTVGESMSTVKGRARPVDHPKRWRRKSWCELGENGVRTVLGCG
                EGAIEVGGDATKPELVAPRRKEVLTTAKLGRSSATAAERERESNDGSSFRVECGWGR
                GGGLESGELVVSNPVAFNGPGKICGRISGQFKRERESGKWREREERTRVVNPPFQAG
                GNGGAMVRRGQQERREVGDDRWVRMLASVVCGAVCVQVGRPSKRGEADQADGRRERNG
                MGHIITL"
gene            complement(25948..26166)
                /gene="OJ1123_G09.8"
CDS             complement(25948..26166)
                /gene="OJ1123_G09.8"
                /note="hypothetical protein"
                /codon_start=1
                /protein_id="BAC00739.1"
                /db_xref="GI:21328162"
                /translation="MPFILVGKIKKTKLDKNAMPRGRFLRSSLVLPPGFCPKKQFCSP
                TSLNATVRNTIAVKKECTKSSGRSLSSL"
LTR             26726..27024
                /note="5' LTR"
gene            27071..31705
                /gene="OJ1123_G09.9"
misc_feature    27071..31705
                /gene="OJ1123_G09.9"
                /note="probably inactive due to stop codon(s) in CDS
                pseudogene, gypsy/Ty-3 retroelement polyprotein"
                /pseudo
LTR             31718..32012
                /note="3' LTR"
gene            complement(join(33051..33520,33601..33713,33765..33994))
                /gene="OJ1123_G09.10"
CDS             complement(join(33051..33520,33601..33713,33765..33994))
                /gene="OJ1123_G09.10"
                /note="hypothetical protein"
                /codon_start=1
                /protein_id="BAC00740.1"
                /db_xref="GI:21328163"
                /translation="MAAARWTKSGRARPGGTMDLDHRWSTGERSSASQLRGGAWPDD
                HRQRDRERHGKRQNGRRGGGGGSPHRAPRRPPATTGGRGRVAEHLLDLAKGMEG
                LGREEVRGGGRCAVGCGERGGCVGLGCGRERWQMVAGGGEHSDDLNFAVKMVIWLC
                FWGGGVDAGVRQGTVMPMMQEARHGDGGSCGSARPEKVVRVCGGGGVSMLRKWWPRC
                GTRRRSRRWWHSAATAMTAMACDRSMLASGGEALGVGSRGEREGSLGEEGK"
gene            complement(join(34860..34963,35445..35480,36798..36852,
                36971..37123,38978..39065,39351..39376))
                /gene="OJ1123_G09.11"
CDS             complement(join(34860..34963,35445..35480,36798..36852,
                36971..37123,38978..39065,39351..39376))
                /gene="OJ1123_G09.11"
                /note="hypothetical protein"
                /codon_start=1
                /protein_id="BAC00741.1"
                /db_xref="GI:21328164"
                /translation="MVPSAVNLSWPISGHKDARRSLLNPPSKALSPMAQLQGWQPDL
                RSKPLVIDSTYVGTRRNSCLLFCYNQVLVLDFWSELCKFELDCLIPTLHSSHRRPSQFA
                SSVHEAQKHYTSERGKNKLGSSLVHRGESRSWTTGHLGRPNSITPSLTIGA"
gene            join(44553..44618,45251..45275,46869..46927,46990..47209,
```

```
                    47387..47664
                    /gene="OJ1123_G09.12"
CDS                 join(44553..44618,45251..45275,46869..46927,46990..47209,
                    47387..47664)

Query Match     7.2%;  Score 192.4;  DB 8;  Length 107491;
Best Local Similarity  67.2%;  Pred. No. 2.7e-36;
Matches  328;  Conservative   0;  Mismatches  136;  Indels   24;  Gaps   3;

Qy   1232  GAGCTCATCATTTATATAAGCATGGTGATCACCAAAATTACTCCTTTGAGCATGTTTATA  1291
           |||||||  ||||      ||||| |||| |  ||||||||   |||    ||||||||||||
Db 100032  GGGCTCATCCGTATATCCTAAGCATGGTGCAACAAACAAAGGTAACACCATTGAATATATG 100091

Qy   1292  TTTAGGAGGACGTTTTACCTGTTGAGGTAAATCTGAACGGCTAATAAATCGGCTAAGCAAA 1351
              |  ||||| |  |||   ||   |||| | ||||   || ||||  |  | |||||||||||
Db 100092  GGCAGAGGGCCATTTTGCATTGTGAGGTGAATCTGGAATCCTCCAGACAAA          100151

Qy   1352  ATAATTTATCACCTGTTGATTCTAACAATTTGATGATGGACAATATTGATGAGGTGACTG  1411
           |||||  ||||| |||| | ||| |||||||||| ||| ||| ||||||| |||||||||
Db 100152  ATAATTGTCGGCAGTAGATTATCACAACTAATGATGACGGAATAGATGAGGTTAGTG      100211

Qy   1412  ACAAATGATTGAAGGCTTTAAAGGAGGATTGAGAGGATAAT----------CT          1455
           | |||||||||||  ||||  || |||||||||||  ||||          ||
Db 100212  ATGAAAGATTGAAAGCTTTGAGAGAGATTGAGAGAGATAAATTGAGAGTAGCTAAAGCTT 100271

Qy   1456  ACAATAAAAATGTAAAGAAGAAAAGCATTCAAAGTGTGAGATCTCGGTGTCGGAAGACTATTT 1515
           ||||||||||  ||||| | |||| | ||| |||| |||| | ||||||| ||||| |   |||
Db 100272  ACAATAAAAAGGTGAAGGAAAAATGTTTCAAATTGTTTAGTGTGAAAACAAATTC       100331

Qy   1516  TGCCTCTTCGGGGGTAAAAGACAACAAGTTTAGTAACGTGCCTCAAAATTGGGAGGGCCCA 1575
           |||||  || |||  ||||||| | ||  ||||| |  ||   |||||||||||| |||||
Db 100332  TGCCTGTT-GGTTCAAAAGATAATAAAATTGGAAAATGGTCTCCAAATTGGGAAGGGCCA 100390

Qy   1576  TGCAAGATTGTTAAAGTAAT------TGTTTTGGATTGACGGAGGCCATTCAAGGTGAT  1628
              | | |||| |  |||  || ||      ||    | |  ||||| |||| | |||||||
Db 100391  TATAGAATTGTAGAAAATAGTCCCCGGAATTCTTATTTTGTGCAAAGTCACGAGGATAT  100450

Qy   1629  CATCTACCTAGAGCTCTCAATGGGAGGTGCTCGAAGACATATTACCCATGTGTATGGCAA  1688
           |||| ||||||||||    ||||||| |  |  ||||||| ||| ||| ||||| |||||
Db 100451  AAGTTGCCTAGAGCTTTTTAATGGGAGATATTTAAAAGGTACTATCCAAGTGTTTGGCAA 100510

Qy   1689  GATGTTTA  1696
           |||| |||
Db 100511  GATGCTTA  100518


RESULT 3
AP004767
LOCUS       AP004767             113514 bp    DNA     linear   PLN 31-AUG-2004
DEFINITION  Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
            BAC clone:OJ1460_H08.
ACCESSION   AP004767 BA000010
VERSION     AP004767.2  GI:20086415
KEYWORDS    .
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 113514)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC
            clone.OJ1460_H08
  JOURNAL   Published Only in Database (2002)
REFERENCE   2  (bases 1 to 113514)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-2002) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     On Apr 8, 2002 this sequence version replaced gi:18844983.
```

```
            Genes were predicted from the integrated results of the following:
            GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor
            (October 1998 version). The genomic sequence was searched against
            NCBI NonRedundant Protein database, nr
            (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
            RGP. Protein homologies of the coding regions were searched against
            NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent
            the identified cDNA sequences using BLASTN 2.0 with the
            corresponding DDBJ accession no. and RGP clone ID.
            A gene with identity or significant homology to a protein is
            classified based on the protein name to indicate the homology level
            such as same name, 'putative-' and '-like protein'. A gene without
            significant homology to any protein but with EST homology (covering
            almost the entire length of partial sequence) is classified as an
            'unknown' protein. A gene predicted with a gene prediction program
            is classified as a 'hypothetical' protein.
            The orientation of the sequence is from -21M13 to M13rev of the BAC
            clone. This sequence of OJ1460_H08 clone has an overlap with
            OJ1123_G09 (DDBJ: AP004613) clone at 5' end and with B1109A06
            (DDBJ: AP004610) at 3' end. The sequence was generated by combining
            Monsanto and RGP-Japan sequencing data. Detailed information on
            overlap and assembly quality together with annotation of this entry
            is available at
            http://rgp.dna.affrc.go.jp/GenomeSeq.html.
FEATURES             Location/Qualifiers
     source          1..113514
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="1"
                     /clone="OJ1460_H08"
     gene            complement(join(778..811,2823..3109))
                     /gene="OJ1460_H08.1"
     CDS             complement(join(778..811,2823..3109))
                     /gene="OJ1460_H08.1"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAB92937.1"
                     /db_xref="GI:20805271"
                     /translation="MSVGWCEADSAGDSSATTKGGANGFRRPSTCGGAARRSRGERCG
                     MATPWPSSRRFGALAIVGDSGQFIEWRSVQSVLGFDFANKVGERSFGMACPRTSHYIS
                     KLFK"
     LTR             5635..5969
                     /note="5' LTR"
     gene            6334..13592
                     /gene="OJ1460_H08.2"
     misc_feature    6334..13592
                     /gene="OJ1460_H08.2"
                     /note="probably inactive due to frameshift(s) in CDS
                     probably inactive due to no initiation codon in CDS
                     probably inactive due to stop codon(s) in CDS
                     pseudogene, putative retrotransposon Cinful prpol"
                     /pseudo
     LTR             13902..14237
                     /note="3' LTR"
     gene            complement(join(15314..15491,19080..19200,20389..20524,
                     21054..21209))
                     /gene="OJ1460_H08.3"
     CDS             complement(join(15314..15491,19080..19200,20389..20524,
                     21054..21209))
                     /gene="OJ1460_H08.3"
                     /note="hypothetical protein
                     similar to Arabidopsis thaliana chromosome 3, At3g23610"
                     /codon_start=1
                     /protein_id="BAB92938.1"
                     /db_xref="GI:20805272"
                     /translation="MSMPEAGRSERDGREQDDYEQQQARVLMALMQGFCAARYRKAD
                     NIPCPIVQGLYLGSVGAAMNKDALKSLNITHILIVARSLNPAFAAEFNYKKIEVLDSP
                     DIDLAKHFDECFSFIDESISSGGNVLVHCFAGRSRSVTTIVAYLMKKHQMSLENALSL
                     VRSKRPQVAPNEGFMSQLENFEKSMQGKIQIAFFTK"
     gene            join(24874..25186,26967..27016)
                     /gene="OJ1460_H08.4"
```

```
CDS             join(24874..25186,26967..27016)
                /gene="OJ1460_H08.4"
                /note="hypothetical protein"
                /codon_start=1
                /protein_id="BAB92939.1"
                /db_xref="GI:20805273"
                /translation="MTLLLVDGGDWVAAREARDAARPAMRAARRSGGRDGKRRLEAV
                EAAVFRVCFRVVVSAAGWGKRGNSRACPCGAAGGASGVARRWLGRRWLSGGGGWQW
                RLALEISIVLPVQVWSRA"

gene            complement(join(42130..42140,42359..42465,44531..44673))
                /gene="OJ1460_H08.5"
CDS             complement(join(42130..42140,42359..42465,44531..44673))
                /gene="OJ1460_H08.5"
                /note="contains ESTs AU082610(C30277),AU062714(C30277)"
                unknown protein"
                /codon_start=1
                /protein_id="BAB92940.1"
                /db_xref="GI:20805274"
                /translation="MFSRHRLRLIVLLSISPQEPRRMVLRVWRSATVFFTSGCRRHCRF
                RLLRQRKLWIFLVRVRIALARALFMNPTILLLDEPTNHLDMA"

gene            53007..53915
                /gene="OJ1460_H08.6"
CDS             53007..53915
                /gene="OJ1460_H08.6"
                /note="hypothetical protein"
                /codon_start=1
                /protein_id="BAB92941.1"
                /db_xref="GI:20805275"
                /translation="MAVTRVAVRGGPGPRTAEGGSAPRGADVAPTWTRGGHVARHAG
                GRTDGPDLPEGRSDGGGERTGAHGRLKPARRPWRTAALPRGGGGRRRRPTASGGGAT
                ARAALAAAKWLGGGAAAKRCGRERERREEGEVSHRRRLJTASDRWRCRRDELLNH
                LANTVKGVREGEEDGRGEDAGRRRRPWRWRTARSGFPORRNPTGEWGWPARRRRTRWR
                LRRGSVQRLETASGGGSGGGGGKEGDGARAMEGTGTDGCKTKKRGSGNGEPLYRPTGV
                DRGRGRDRDFASGVGRGRGBEREVGFKN"

LTR             complement(54043..54476)
                /note="3' LTR"
gene            complement(54488..58879)
                /gene="OJ1460_H08.7"
misc_feature    complement(54488..58879)
                /gene="OJ1460_H08.7"
                /note="probably inactive due to stop codon(s) in CDS
                probably inactive due to frameshift(s) in CDS
                pseudogene, gag-pol"
                /pseudo

gene            complement(join(62255..62275,62449..63042))
                /gene="OJ1460_H08.8"
CDS             complement(join(62255..62275,62449..63042))
                /gene="OJ1460_H08.8"
                /note="hypothetical protein
                similar to Oryza sativa chromosome 3, OSJNBb0031G04.25"
                /codon_start=1
                /protein_id="BAB92942.1"
                /db_xref="GI:20805276"
                /translation="MADHQPNHRGDGMDDFVAELARMTLVVGYPNAPEYSTVPPLSGE
                LPHRVRLEVHCYLGTSLANMVEASGGTADHAYQEAAYLMMARLRDHHSYIFCNTAYR
                FHPRRASGNDISTFRPTVGENDTTFGHMCAVMRGMDRMHSDLHKASKALNDGKLVRII
                ALKDEIARLKKENAQLKGLPAPGGVRIRTTPRKSTTTPIFLGII"

LTR             complement(66254..66702)
                /note="5' LTR"
gene            join(76774..76835,77289..77457)
                /gene="OJ1460_H08.9"
CDS             join(76774..76835,77289..77457)
                /gene="OJ1460_H08.9"
                /note="hypothetical protein"
                /codon_start=1
                /protein_id="BAB92943.1"
                /db_xref="GI:20805277"
                /translation="MGVLTGRPRPFAGSAGGANDFLAILALKPGMVLTGTNIPTPG
                VSTPAPARDESQEALIIPTGRGEACDRPPVAT"

gene            78396..81540
                /gene="OJ1460_H08.10"
misc_feature    78396..81540
                /gene="OJ1460_H08.10"
```

```
                /note="probably inactive due to frameshift(s) in CDS
                probably inactive due to stop codon(s) in CDS
                pseudogene, similar to Oryza sativa unknown protein,
                OJ1208D02.7"
                /pseudo
gene            complement(82160..87456)
                /gene="OJ1460_H08.11"
misc_feature    complement(82160..87456)
                /gene="OJ1460_H08.11"
                /note="probably inactive due to frameshift(s) in CDS
                probably inactive due to no initiation codon in CDS
                probably inactive due to stop codon(s) in CDS
                pseudogene, gag-pol precursor"
                /pseudo
gene            join(93167..93198,93628..93733,94179..94237,94285..94351,
                94634..94659,95457..95583,97893..97925)
                /gene="OJ1460_H08.12"
CDS             join(93167..93198,93628..93733,94179..94237,94285..94351,
                94634..94659,95457..95583,97893..97925)
                /gene="OJ1460_H08.12"
                /note="contains ESTs C72693(E2071),AU101165(E2071)
                unknown protein"
                /codon_start=1
                /protein_id="BAB92944.1"
                /db_xref="GI:20805278"
                /translation="MPRHPKTQRTSVVLEDHINDSDQSSTIAGCSACCDHAMTSENTE
                DQRFQSIANNMRLVDLKLRMVWFDTFGLHTRIFHVLQLSIIHLQRRKGSEENLRKYLR
                PCLVSQESRSWSHANMHHANVVLSFQIRPCCIDYIQAGYRKIFQP"
LTR             complement(103810..104012)
```

```
Query Match          7.2%;  Score 192.4;  DB 8;  Length 113514;
Best Local Similarity 67.2%;  Pred. No. 2.7e-36;
Matches  328;  Conservative  0;  Mismatches 136;  Indels 24;  Gaps  3;

Qy      1232 GAGCTCATCATTTATATAAGGCATGGTGATACCAAAATTACTCCTTTGAGCATGTTTATA 1291
             ||||||||||||   |||||   |||||  ||||||||| || |||||||||| ||| |
Db     13105 GGGCTCATCGTATATCTAAGCATGGTGCAACAAAAGTAACACCATTTGAATTAGTATATG 13164

Qy      1292 TTTAGGAGGACGTTTACCTGTTGAGGTAAATCTGAACGCTAATAAATCGGCTAAGCAAA 1351
             |||||| ||  |||| |||||||||||||||  ||| ||  |||| |||| | ||| |||
Db     13165 GGCAAGAGGCCATTTTGCATGTTGAGGTAATCTGACGCCCTAGATTGCCAGACAAA 13224

Qy      1352 ATAATTATCACCTGTTGATTCTAACAATTTGATGATGACAATATTGATGAGGTGACTG 1411
             ||||||   || |||||| |||  ||||||||   || |||| ||||||| ||||||||
Db     13225 ATAATTTGTCGGCAGTAGATTATCACCAACTTAATCATGAGATTGAGGAGGTTAGTG 13284

Qy      1412 ACAAATGATTGAAGGCTTTAAAGGAGATTGAGAACGGATAAT------------CT 1455
             |||| ||||| ||||||||||| |||| ||||||| |||||||            ||
Db     13285 ATGAAAGATTGAAAGCTTTGAGAGAGATTGAGAGAAGATAAATTGAGAGTAGCTAAAGCTT 13344

Qy      1456 ACAATAAAAATGTAAAGAAGAAGAAGCATTCAAAGTTGAGATCTCGGTGTGGAAGACTATT 1515
             |||||||||||||||||| ||||||||  ||||||||||||   |||||||| | ||||||
Db     13345 ACAATAAAAAGGTGAAGGAGAAAATCGTTTCAAATTGGAGATTTAGTGTGTGAAAACAATTC 13404

Qy      1516 TGCCTCTTGGGGTAAAAGACAACAAACAAGTTTAGTAAGTGGCCTCAAAAATTGGGAGGGCCCA 1575
             |||||||  ||||||||||| |||||| |||| ||| |||| ||||||||||||||| ||||||
Db     13405 TGCCTGTT-GGTTCAAAGATAAAAAGATAAAATTTGGGAAATGGCTCCCAAATTGGGAAGGGCCA 13463

Qy      1576 TGCAAGATTGTTAAAGTAAT------------TGTTTTGGATTGACGGAGGCATTTCAAGGTGAT 1628
             ||||||||| ||||||| ||            ||||| || ||  |||| | || ||| |||| |
Db     13464 TATAGAATTGTAGAAATAGTCCCCGGGAATTCTTATTTTGTGCAAAGTCTTACGAGGATAT 13523

Qy      1629 CATCTACCTAGAGCTCTCAATGGGAGGTGCTCGAAGACATATTACCCATGTGTAATGCGCAA 1688
             ||||||||||||||||||| || ||||||||||| ||||| ||||| |||| |||||| |||
Db     13524 AAGTTGCCTAGAGCTTTTAATGGGAGATATTTAAAAAGGTACTATCCAAGTTGTTTGGCAA 13583

Qy      1689 GATGTTTA 1696
             |||| |||
Db     13584 GATGCTTA 13591
```

```
RESULT 4
AP004574
```

```
LOCUS       AP004574              144596 bp    DNA     linear   PLN 03-DEC-2003
DEFINITION  Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 8,
            PAC clone:P0702G08.
ACCESSION   AP004574
VERSION     AP004574.3  GI:38636932
KEYWORDS    .
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 8, PAC
            clone:P0702G08
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 144596)
  AUTHORS   Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-DEC-2001) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     On Dec 2, 2003 this sequence version replaced gi:33235572.
            Genes were predicted from the integrated results of the following:
            GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH
            (http://www.softberry.com/), GeneMark.hmm
            (http://opal.biology.gatech.edu/GeneMark/), GlimmerM
            (http://www.tigr.org/tdb/glimmerm/glmr_form.html), RiceHMM
            (http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor
            (http://bioinformatics.iastate.edu/cgi-bin/sp/sp.cgi), sim4
            (http://globin.cse.psu.edu/html/docs/sim4.html), gap2
            (http://www.tigr.org/software/glimmerm/), BLASTN and BLASTX. The
            genomic sequence was searched against NCBI NonRedundant Protein
            database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA
            sequence database at RGP or DDBJ. Protein homologies of the coding
            regions were searched against NCBI NonRedundant Protein database
            with BLASTP. ESTs represent the identified cDNA sequences using
            BLASTN with the corresponding DDBJ accession no. and RGP clone ID.
            Full-length cDNAs represent the identified cDNA sequences using
            BLASTN with the corresponding DDBJ accession no.
            A gene with identity or significant homology to a protein is
            classified based on the protein name to indicate the homology level
            such as same name, 'putative-' and '-like protein'. A gene without
            significant homology to any protein but with full-length cDNA or
            EST homology (covering almost the entire length of partial
            sequence) is classified as an 'unknown' protein. A gene predicted
            by two or more gene prediction programs is classified as a
            'hypothetical' protein according to IRGSP standard. A gene
            predicted by a single gene prediction program is also classified as
            a probable 'hypothetical' protein and is included as a
            miscellaneous feature of the sequence.
            The orientation of the sequence is from SP6 to T7 of the PAC clone.
            This sequence of P0702G08 clone has an overlap with P0683E12 (DDBJ:
            AP004706) clone at 5' end and with OSJNBa0062G05 (DDBJ: AP005491)
            at 3' end. Detailed information on overlap and assembly quality
            together with annotation of this entry is available at
            http://rgp.dna.affrc.go.jp/GenomeSeq.html.
FEATURES             Location/Qualifiers
     source          1..144596
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="8"
                     /clone="P0702G08"
     gene            complement(2836..3303)
                     /gene="P0702G08.1"
     misc_feature    complement(2836..3303)
                     /gene="P0702G08.1"
                     /note="putative gypsy-type retrotransposon orf5"
     gene            complement(4240..7583)
```

```
     mRNA            complement(join(4240..4693,4959..5060,5167..5244,
                     5348..5441,5561..5739,5814..5930,6295..7048,7154..7583))
                     /gene="P0702G08.2"
                     /note="supported by full-length cDNA(s): AK100576"
     CDS             complement(join(4562..4693,4959..5060,5167..5244,
                     5348..5441,5561..5739,5814..5930,6295..6852))
                     /gene="P0702G08.2"
                     /note="contains EST(s): AU085915(E2817),C73047(E2817),
                     contains full-length cDNA(s): AK100576"
                     /codon_start=1
                     /product="GTPase activator protein-like"
                     /protein_id="BAD03195.1"
                     /db_xref="GI:38636933"
                     /translation="MFGTQVRGVVEDDFNPRRRRTLWNTPAASMPTTHTIRYSTPRK
                     GHANAHQLLAKFEDLYGFWVEGNVDDVNVLNDVRERMREQGRVWWALEASKGANWYLQ
                     PRISSNGGSEGVISVTSLKLSVLTNTITLRRLIRKGVPPVLRPKVWLSVSGAAKKRST
                     VPETYYDELIRAIEGKTTPATLQIDHDLPRTFPCHSWLNSEEGQASLRRVLVGYSFRD
                     SEVGYCQGLNYVAALLLVMKTEEDAFWMLAVLENVLVNDCYTDNLSGCHVEQRVFK
                     DLLAKKCPRIAAHLEAMGFDVSLVATEWFLCLFSKTLPSETTLRVWDVLFNEGAKVLF
                     HVALAIFKQREDDLLRIQHIGDVIDILQTTAHHLYDPDELLTFAPDKIGSMTTNTITK
                     ERKKQETVVMAELDQRIRRLSSLKMDG"
     gene            complement(10029..10368)
                     /gene="P0702G08.3"
     mRNA            complement(join(<10029..10235,10267..>10368))
                     /gene="P0702G08.3"
                     /note="start and end point are not identified"
     CDS             complement(join(10029..10235,10267..10368))
                     /gene="P0702G08.3"
                     /note="predicted by GENSCAN etc."
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="BAD03196.1"
                     /db_xref="GI:38636934"
                     /translation="MLCRKTVLLRPPSLCTAALIYRREEPTRRRSASFRRRFASSTTA
                     RSPSAAATYPPPAPPLRRLQHHEDVRRRYISSAATAPPNRAIIARWTGKGGERKG
                     "
     gene            complement(join(16418..16454,17425..17569,17753..17807))
                     /gene="P0702G08.4"
     misc_feature    complement(join(16418..16454,17425..17569,17753..17807))
                     /gene="P0702G08.4"
                     /note="hypothetical ORF
                     predicted by RiceHMM
                     this category is not included in IRGSP standard"
     gene            complement(join(20050..20162,20764..20874,23940..24134,
                     24663..24675))
                     /gene="P0702G08.5"
     misc_feature    complement(join(20050..20162,20764..20874,23940..24134,
                     24663..24675))
                     /gene="P0702G08.5"
                     /note="hypothetical ORF
                     predicted by GENSCAN
                     this category is not included in IRGSP standard"
     gene            complement(27170..27709)
                     /gene="P0702G08.6"
     misc_feature    complement(27170..27709)
                     /gene="P0702G08.6"
                     /note="retroelement-like"
     gene            complement(join(29847..30222,30334..30602,31708..31746))
                     /gene="P0702G08.7"
     misc_feature    complement(join(29847..30222,30334..30602,31708..31746))
                     /gene="P0702G08.7"
                     /note="start and end point are not identified"
     gene            complement(join(35762..38932))
                     /gene="P0702G08.8"
                     /note="polyprotein-like"
     mRNA            join(<35762..36165,37420..38101,38186..>38932)
                     /gene="P0702G08.8"
     CDS             join(35762..36165,37420..38101,38186..38932)
                     /gene="P0702G08.8"
                     /note="similar to Arabidopsis thaliana chromosome 3,
                     At3g50120"
                     /codon_start=1
```

```
Db    18128  AAAGCCATTAATCCTTTTGTAATGATACTTGCATACAAATTATTCCAAGCCATACATGTA  18187
                     |||||||||||||||||||||||| ||||||||||  || |||||||||||||||||
Qy      769  CATGTGAACTTGAAGTGCTACCACGCCATCTTGGGTGCTTGGGTGGTTGAGAAATGGAACAGAGATCA  828

Db    18188  CTTGTGAACCTTAAAGTGTCTA-CATGTCATCTTGAATGATTGAAGAATGAAACGCCTT  18246
                     ||||||||| |||||||||  |||  || ||| ||||  ||||||  |||||||||||
Qy      829  GAATCCAGTCATCTTTACCTCTCTTGCTTAGTAGAACTTGGGTTTTGTAAGGTTTTCAAA  888

Db    18247  GACTCCAAGTTACCTTCCATTCCTTTTAATGAATTCCTTGAGGTTTTCTTTTCTCATAAA  18306
                     || || |||||||| |||| || || ||||||||||||||||  |||||||||||||||
Qy      889  TTTAAAACATAGTCTTGCTCCTCAAATGATTCCTCATATAGCTCAATGTGTATGGTTTC  948

Db    18307  AGTATAACTTTACTTTACTCCTTAATTGATTCTCCGGATCGCTCAATGTGTATGATTCC  18366
                     |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Qy      949  TCACCAAGGCAATG-TTTTGCCTCCTTTCATCCTACACTTCCTATATATTTCTTTTGTGGAGCT  1007

Db    18367  TCACCAAGGCATTGCATTTTTGCATTTTCTTATCCTACACTTCTAATAGACTTATGTAGAGCT  18426
                     ||||||||||  |||| ||| |||||||  || |||||||||| || |||| ||||||||||
Qy      1008 TAGGGTAGGGAATGAAAAGGAAGCCATACTTGCATTGCATATGTTACTAAGTCAAAAA  1064

Db    18427  CAAGTAGGGAATAAGAAGTGA--CATACTTGCATCGCATATATTGTAAAGTCAAAA  18481
```

```
RESULT 5
AP004706
LOCUS        AP004706        165766 bp        DNA        linear    PLN 03-DEC-2003
DEFINITION   Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 8,
             PAC clone:P0683E12.
ACCESSION    AP004706
VERSION      AP004706.3  GI:3863711B
KEYWORDS     .
SOURCE       Oryza sativa (japonica cultivar-group)
  ORGANISM   Oryza sativa (japonica cultivar-group)
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
             Ehrhartoideae; Oryzeae; Oryza.
REFERENCE    1
  AUTHORS    Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE      Oryza sativa nipponbare(GA3) genomic DNA, chromosome 8, PAC
             clone:P0683E12
  JOURNAL    Published Only in Database (2002)
REFERENCE    2  (bases 1 to 165766)
  AUTHORS    Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE      Direct Submission
  JOURNAL    Submitted (30-JAN-2002) Takuji Sasaki, National Institute of
             Agrobiological Sciences, Rice Genome Research Program; Kannondai
             2-1-2, Tsukuba, Ibaraki 305-8602, Japan
             (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
             Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT      On Dec 2, 2003 this sequence version replaced gi:34013504.
             Genes were predicted from the integrated results of the following:
             GENESCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH
             (http://www.softberry.com/), GeneMark.hmm
             (http://opal.biology.gatech.edu/GeneMark/), GlimmerM
             (http://www.tigr.org/tdb/glimmerm/glmr_form.html), RiceHMM
             (http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor
             (http://bioinformatics.iastate.edu/cgi-bin/sp_cgi), sim4
             (http://globin.cse.psu.edu/html/docs/sim4.html), gap2
             (http://www.tigr.org/software/glimmerm/), BLASTN and BLASTX. The
             genomic sequence was searched against NCBI NonRedundant Protein
             database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA
             sequence database at RGP or DDBJ. Protein homologies of the coding
             regions were searched against NCBI NonRedundant Protein database
             with BLASTP. ESTs represent the identified cDNA sequences using
             BLASTN with the corresponding DDBJ accession no. and RGP clone ID.
             Full-length cDNAs represent the identified cDNA sequences using
             BLASTN with the corresponding DDBJ accession no.
             A gene with identity or significant homology to a protein is
             classified based on the protein name to indicate the homology level
             such as same name, 'putative-' and '-like protein'. A gene without
             significant homology to any protein but with full-length cDNA or
             EST homology (covering almost the entire length of partial
             sequence) is classified as an 'unknown' protein. A gene predicted
```

```
             /product="hypothetical protein"
             /protein_id="BAD03197.1"
             /db_xref="GI:38616935"
             /translation="MMSPAAAASASPPTGSSGRGEGVARHCCLSSRHLPPLPPPPAAA
             AIVASPLAGFGGGEGAPAADRCRLSSCRRPLPPPPTSAANSHHCRCHLSSRCPRLPP
             LRPPPSAAGEKVREEAQKRERVRERRRDARANIKKSHOMEGEEQORVKTTSQWVSH
             VEHLLDAGNPSEEEARRWRRQSVRYRVPAYIKHRTSYGAPQLVSLGPFHHGVPELRPA
             EEHKERALLHLLRRGGGGDGRRLRLGSLVASMEEVVVELQDAYQGLGATKWRDDRDG
             FLKMMVLDGCFLLEWMRAAAEAAQGDGEGVGGGYAANDPVFSRHGELYVFPYVRRD
             MLMIENQLPLLVLQRIVAFVHGGAAPEASDDAINNMVLSFVSMIPDDPAMRGGGGLAL
             HPLDVCHRSLLHGSPPRCHTGRREEFVPSATELDQAGVRFRPSRTRSLHDISFRHGA
             LRIPRLAVDDTTEHKLFSLMAFEQLHGAGANEVTAYVFFMDNVIKSGDDARLLGASGV
             VSNGLGSDEAVAEMFNRLASEAELDRRSALHGVHGEVNAYREKRWNQWRASLVRNHAG
             NPWAIVSLVVAFVLLVLTVLQVTVYTVLPYYQEQAAAGEAALREL"
gene         39453..39734
             /gene="P0702G08.9"
mRNA         <39453..>39734
             /gene="P0702G08.9"
             /note="start and end point are not identified"
CDS          39453..39734
             /gene="P0702G08.9"
             /note="predicted by FGENESH etc."
             /codon_start=1
             /product="hypothetical protein"
             /protein_id="BAD03198.1"
             /db_xref="GI:38616936"
             /translation="MPPPPPPPRRTAAATAAADNDNDNVDYAAEMEADAAGNGSARS
             AAAAVARAVVYDGVDPFEGMEFDNEEPLTRTSWCRHSRSSHRGWTRRR"
gene         complement(40372..42402)
             /gene="P0702G08.10"
mRNA         complement(join(40372..41025,41120..41428,41561..41769,
             41912..42402))
             /gene="P0702G08.10"
             /note="supported by full-length cDNA(s): AK068483"
CDS          complement(join(40816..41025,41120..41428))
             /gene="P0702G08.10"
             /note="contains EST(s): AU093746(E0094),C71705(E0094)
             contains full-length cDNA(s): AK068483
             similar to Sorghum bicolor chromosome 4, SB18C08.27"
             /codon_start=1
             /product="unknown protein"
             /protein_id="BAD03199.1"
             /db_xref="GI:38636937"
             /translation="MAGGRTYWDKALTKIFLDLCIAEKTKRNHNKKGRTNIGWQNLYR
             NFREQSGKNYDSKQLONKFSTFKXQYKLWKSLKNKSGGGWDNNSSTIRCDDDWEDRI
             EENRDARQFRGKPLEHEDELTHFGCMDTEEGTMLCVGGIGKRTPSSGSGDDNLTPMSN
             DNVGLSSAGRVA"
gene         complement(43789..52522)
             /gene="P0702G08.11"
mRNA         complement(join(43789..44169,44514..44582,44852..44899,
             45729..45806,45915..46028,47061..47252,47344..47601,
             47604..48034,48157..48303,48368..48481,48636..48636,
             48805..48861,49488..49532,49666..49761,49857..49932,
             50703..50841,51108..51171,51279..51352,51670..51748,
```

```
Query Match      7.0%;    Score 188.6;    DB 8;   Length 144596;
Best Local Similarity   66.9%;    Pred. No. 2.5e-35;
Matches  359;  Conservative   0;  Mismatches 169;  Indels  9;  Gaps   6;

Qy    533  ATAAATACCAAAACTCATGGAGCTTGCTAGTATAAAACTCTAATTCTAAGTTTG---GT  589
                  |||  ||||||  |||  ||||  ||| || |||| |||||||||||||||||||||
Db  17949  ATATTTCACCAATACTTGTGAAAATAATTAGTAATAACCACTACCACTCAAGTTTGATTTT  18008

Qy    590  GTTTATTTGAGTCGATTTTGTGTGAAAGTTGGTCGGTTAGAAATAGGAGTTAAGGACCGCC  649
                  |||||  ||||||| || ||||||||||  ||||||||||||||||||||||||||||
Db  18009  TTTTATTTCAATTGATTTATGCAAGGTTTGACAAGGTTTGACCATCGAATTCGGATTCGGTTGCCAT  18068

Qy    650  AACAAGATCCCCACCACTTAGCCCTTTGCTCATCCTCCGAGTAAAGTTCAAGGACTAAGGT  709
                  ||||||| |||||||||||||||| ||| ||||||||||||||||||||||||||||||
Db  18069  CAACAAAATCTCCACACTTAGCCATTA-CTGGTCCCTGAGTAATGATCATGAGTTACACAC  18127

Qy    710  GGAACATCCTCAAATGGT-ACGATGCCTAGATAGTTATTCCAAGCCTCACCTATA  768
                  | ||| ||| ||  |  ||| ||  |||||  || || ||||  ||   |  ||||
```

by two or more gene prediction programs is classified as a
'hypothetical' protein according to IRGSP standard. A gene
predicted by a single gene prediction program is also classified as
a probable 'hypothetical' protein and is included as a
miscellaneous feature of the sequence.
The orientation of the sequence is from T7 to SP6 of the PAC clone.
This sequence of P0683E12 clone has an overlap with P0689D06 (DDBJ:
AP004621) clone at 5' end and with P0702G08 (DDBJ: AP004574) at 3'
end. Detailed information on overlap and assembly quality together
with annotation of this entry is available at
http://rgp.dna.affrc.go.jp/GenomeSeq.html.

```
FEATURES             Location/Qualifiers
     source          1..165766
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="8"
                     /clone="P0683E12"
     gene            complement(2914..4170)
                     /gene="P0683E12.1"
     mRNA            complement(<2914..>4170)
                     /gene="P0683E12.1"
                     /note="start and end point are not identified"
     CDS             complement(2914..4170)
                     /gene="P0683E12.1"
                     /note="similar to Oryza sativa chromosome 4,
                     OSJNBb0072M01.3"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="BAD03373.1"
                     /db_xref="GI:38637119"
                     /translation="MDEHLAYHLPDLALEIVLSHLQSLADRASFRGVCRQWAAVWRDQ
                     WPRTPPMPWLAAPGHCVDLSDASVHRRVPLPSGVDVDGIVCCGSLGNWIALTPKRRWR
                     PRHQVRPLLLNPFSGASVQLPILTPAAFRGGGDDINVEKIVMSSAPDSDGCVVAAIVM
                     GSYSSTREIVIWRRGQDSCSAPAAAPSNVADAVFHGGDLYVVDKCSQLYVFPNHVFSS
                     DGGGGQELHPVRLEMDLTRTSRFVARVLLECDGRLLMADRHRHGGDAGYHEYRVYALE
                     RDASCGDWRWSPVTRLDGHVLFLGAGCCRALPVTGRDRVKDGNVVFLDDSAEITAVVT
                     VDDRKPLERSALIRRSMDVPASNVLDTFRRRGGGGGGADRPAPPAACMAGRRNQCFG
                     FGGLQDLIVLLKSFVSLQEDGYTHRQ"
     gene            complement(4870..5415)
                     /gene="P0683E12.2"
     mRNA            complement(join(<4870..4876,5264..>5415))
                     /gene="P0683E12.2"
                     /note="start and end point are not identified"
     CDS             complement(join(4870..4876,5264..5415))
                     /gene="P0683E12.2"
                     /note="predicted by GeneMark.hmm etc."
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="BAD03374.1"
                     /db_xref="GI:38637120"
                     /translation="MAAAELEDNASGDWRPSLGTARAATATTVLGGCDGNGCAWGRNV
                     FIWVGVRI"
     gene            6094..13263
                     /gene="P0683E12.3"
     mRNA            join(<6094..6298,9294..9339,10053..10167,11333..11476,
                     11566..11892,12319..12512,13128..>13263)
                     /gene="P0683E12.3"
                     /note="start and end point are not identified"
     CDS             join(6094..6298,9294..9339,10053..10167,11333..11476,
                     11566..11892,12319..12512,13128..13263)
                     /gene="P0683E12.3"
                     /note="predicted by GeneMark.hmm etc."
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="BAD03375.1"
                     /db_xref="GI:38637121"
                     /translation="MARARGLYGLASDNGNGRGQRPTGRHTADNGRASNGWRGRERGQ
                     SSHTGGGCREGARPLDWERRTGAVGIVDCFAPRPACCOIKGRALAGHKQGYMVAAFTGS
                     TVVNHAPNLLTHAFTPILLAQREEGEAAGRCKGRLRRLTRDGDGWRRRTNKADGR
                     LWGNLERRPAEEAKAVEGEVVDDGTARRFGGGEAANRGGGGGLAGGGEVAAADWSSAV
                     AADWLGEVAAAARWRRRIGWRRRRLVAAVAAADWSGGDEVAAADWLVAARWRRRIGW
                     AAASGSIDLSAIDIHSGAAGVSATTPPSPSLAAGQLDAPPPSAAAARPRPTATKRPR
                     RYCSPAAQFPGALTAGGAGQREWRTAEWWRRTAAGVVANGGEVVALTAVPSDA"
     gene            14594..17717
                     /gene="P0683E12.4"
     mRNA            join(<14594..14973,15522..15790,15942..16055,16481..16584,
                     17601..>17717)
                     /gene="P0683E12.4"
                     /note="start and end point are not identified"
     CDS             join(14594..14973,15522..15790,15942..16055,16481..16584,
                     17601..17717)
                     /gene="P0683E12.4"
                     /note="similar to Oryza sativa chromosome
                     10,OSJNBa0061H20.13"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="BAD03376.1"
                     /db_xref="GI:38637122"
                     /translation="MEVTTIAISVCVKTVSISFAFPSLHTMPNHFFVDHLPFCASQPP
                     EHDHANDALEHAPERRRFDIIVRPFSSIPKQAAMPRISPRPPLQDASRTLARNIKNT
                     ENTSFSSSTLIRTHLRRHRRLPLPRWGPVGPVHLPLPARRRAETAAAPTSPPSPG
                     HVNHARACAASPAPTPPSLPLAPAPEMAGIRKSNPASLSSPTHVGRRFPPLPATSGS
                     PPGCAVAVGFAWPRSTPSTPPSRFVAGALIPSRTAAFASPFAPSPPPAAVLRRCR
                     NRAAPSRAMAAGPARVGVMERGGPCDGLCRFRIHLGTRGDCPLPAGDGGET"
     gene            complement(join(18175..18279,18722..18772))
                     /gene="P0683E12.5"
     misc_feature    complement(join(18175..18279,18722..18772))
                     /gene="P0683E12.5"
                     /note="hypothetical ORF
                     predicted by GeneMark.hmm
                     this category is not included in IRGSP standard"
     gene            23449..23936
                     /gene="P0683E12.6"
     mRNA            join(<23449..23467,23530..>23936)
                     /gene="P0683E12.6"
                     /note="start and end point are not identified"
     CDS             join(23449..23467,23530..23936)
                     /gene="P0683E12.6"
                     /note="predicted by GENSCAN etc."
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="BAD03377.1"
                     /db_xref="GI:38637123"
                     /translation="MANQSIFAARRGIIPNTTSHLMKNRNVSIKGNQHLVVYHIQRHN
                     KVEARSYEGLHCKTETEPPWESSRDALREPLWRRLGSTMGRSKRMGWAGCAMAMPSP
                     AVLVPELSLWIPRELFEVNVHGRRSREAGSRQRRVETNS"
     gene            complement(24717..24878)
                     /gene="P0683E12.7"
     misc_feature    complement(24717..24878)
                     /gene="P0683E12.7"
                     /note="hypothetical ORF
                     predicted by GlimmerM
                     this category is not included in IRGSP standard"
     gene            complement(25230..26573)
                     /gene="P0683E12.8"
     mRNA            complement(join(<25230..25246,26420..>26573))
                     /gene="P0683E12.8"
                     /note="start and end point are not identified"
     CDS             complement(join(25230..25246,26420..26573))
                     /gene="P0683E12.8"
                     /note="predicted by GENSCAN etc."
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="BAD03378.1"
                     /db_xref="GI:38637124"
                     /translation="MWPGTLGDVGPEGEVSALLTDFPEGGRVTPILAPRABAPRPLAF
                     LLHVYGLGQPGR"
     gene            27237..30542
                     /gene="P0683E12.9"
     misc_feature    27237..30542
                     /gene="P0683E12.9"
                     /note="probably inactive due to including stop codon(s) in
                     CDS
                     pseudogene, retroelement"
```

```
gene            /pseudo
                complement(join(31189. .33339,33841. .34854))
                /gene="P0683E12.10"
misc_feature    complement(join(31189. .33339,33841. .34854))
                /gene="P0683E12.10"
                /note="putative GAG-POL precursor"

Query Match          7.0%;   Score 188.6;  DB 8;  Length 165766;
Best Local Similarity  66.9%;  Pred. No. 2.6e-35;
Matches 359;  Conservative  0;  Mismatches 169;  Indels  9;  Gaps  6;

Qy    533 ATAAATCACCAAAACTCATGGAGCTTGCTAGTTATAAACTCTAATTCTAAGTTTG---GT 589
          ||| |||||||||||||||||  || ||||||  ||||  |||||||||||||||    ||
Db 160376 ATATTCACCAATACTTGTGAAAATAATTAGTAATAACCACTACCACTAAGTTTGATTTT 160435

Qy    590 GTTTATTTGAGTGGATTTCGTGTGAAAGTTGGTGTGGTTAGAGAATAGGAGTTAAGGACCGCC 649
          ||||| || |||||| || |||||  |    |||||  |||| |||||||||||||||||||
Db 160436 TTTTATTTCAATTGATTTATGCAAGGTTGACCGATCGGATTTGGTACCTTAAAGACAGT 160495

Qy    650 AACAAGATCCCCACACTTAGCCCTTTGCTCATCCTGAGTAAAGTTCAAGGACTAAGGT 709
          || ||||||||||||||||||| || ||||||  ||||| |||||||||||||| |||
Db 160496 CAACAAAATCTCCACACTTAGCCATTA-CTGGTCCCTGAGTAATGATCATGGATTACAAC 160554

Qy    710 GGAACATCCCCTCAAATGGT-ACGATGCCTGCATATAAGTTATTCCAAGGCCTCACCTATA 768
          ||||||||||  |||||||| |  |||   || |||||||||| |||||||||||||||
Db 160555 AAAGCCATTAATCCTTTGTAATGATAACTTGCATACAAATTAATTCCAAGGCCATACATGTA 160614

Qy    769 CATGTGAACTTGAAGTGTCTACCACGCCATCTTGGGTGGTTGAGAAAATGGAACAGATCA 828
          |||||||||||||||||  ||   ||||||| ||||  |||| |||||| |||||||||
Db 160615 CTTGTGAACCTTAAAGTGTCTA-CATGTCATCTTGAATCTTGAATGATTGAAGAATGAAACAGCCTT 160673

Qy    829 GAATCCAGTCATCTTTACCTCTCTTGCTTAGATAAACTTGGGTTTTGTAAGGTTTTCAAA 888
          |||||||||||||||  |||||||||||  ||||| ||||||||| |||||||||||||
Db 160674 GACTCAAGTTACCTCCATTCCTTTTAATGAATTCCTTGAGGTTTTCTTTTTCTTCATAAA 160733

Qy    889 TTTAAAACATAGTCTTGCTCCTCCCAAATGATTCTTCTCATATAGCTCAATGTGTATGGTTTC 948
          || || |  ||| || || |  ||||||||||  ||||||  |||||||||||||||||||||
Db 160734 AGTATAACTTTACTTTACTCCTTAATTGATTCTCTCGGATCGCTCAATGTGTATGATTCC 160793

Qy    949 TCACCAAGGCAATG-TTTTGCCTCTTTCATCCTACTTCTCAATATTTCTTTTGTGGAGCT 1007
          ||||||||||| || |||||| |||||| |||||||| || ||||||| |||||||||||
Db 160794 TCACCAAGGCCATTGCATTTTGCATTTCTTTATCCTTCTTCTCTAAAGACTTATGTAGAGCT 160853

Qy   1008 TAGGGTAGGGAATGAAAAGGAAGCATACTTGCATTGCATATGTTACTAAGTCAAAA 1064
          |||||| ||||||||||  |  || || |||||||||||||  ||||| |||||||
Db 160854 CAAGGTAGGGAATAAGAATGA--CATACTTGCATCGCATATATTGTAAAGTCAAAAA 160908
```

```
RESULT.6
OSJN00028
OSJN00028
LOCUS       OSJN00028               123269 bp    DNA     linear   HTG 11-SEP-2001
DEFINITION  Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 4 clone
            OSJNBa0055E05, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AL606596
VERSION     AL606596.1  GI:15594055
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (japonica cultivar-group)
ORGANISM    Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
```

```
REFERENCE   1
AUTHORS     Fu,G., Wang,S.Y., Ren,S.X., Lv,G., Lin,W., Gu,W.Q., Zhu,G.F.,
            Tu,Y.F., Jia,J., Yin,H.F., Zhang,Y., Cai,Z., Chen,J., Kang,H.,
            Chen,X.Y., Shao,C.Y., Sun,Y., Hu,Q.P., Zhang,X.L., Zhang,W.,
            Wang,F.H., Ding,C.W., Sheng,H.H., Gu,J.L., Chen,S.T., Ni,L.,
            Zhu,F.H., Han,B., Feng,Q., Huang,Y.C., Li,Y., Zhu,J.J., Zhao,Q.,
            Hu,X., Liu,Y.L., Mu,J., Yu,Z., Chen,L., Fan,D.L., Weng,Q.J.,
            Zhang,L., Lu,Y.Q., Yu,S.L., Liu,X.H., Lu,T.T., Zhang,Y.G., Lu,Y.,
            Li,C., Li,T., Zhang,Y., Hu,H., Jia,P.X., Qian,Y.M., Ying,K.,
            Zhou,B., Chen,Z.H., Hao,P., Zhang,L., Wu,M., Zhang,R.Q., Guan,J.P.
            and Hong,G.F.
TITLE       Direct Submission
```

```
JOURNAL     Submitted (28-JUL-2000) Han Bin, National Center for Gene Research,
            Chinese Academy of sciences, 500# Cao Bao Road, Shanghai 200233,
            CHINA. E-mail enquiries: bhan@ncgr.ac.cn. Clone requests:
            bhan@ncgr.ac.cn
REMARK      Oryza sativa japonica (nipponbare) genomic DNA, chromosome 4, BAC
            clone: OSJNBa0055E05.
            Web site: http://www.ncgr.ac.cn
COMMENT     -------------------- Summary Statistics
            Assembly program: phrap
            --------------------
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 1 contigs.Gaps between the contigs
          * are represented as runs of N. The order of the pieces
          * is believed to be correct as given, however the sizes
          * of the gaps between them are based on estimates that have
          * provided by the submittor.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *      1 123269: contig of 123269 bp in length.
```

```
FEATURES             Location/Qualifiers
     source          1..123269
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /variety="Nipponbare"
                     /sub_species="japonica"
                     /db_xref="taxon:39947"
                     /chromosome="4"
                     /clone="OSJNBa0055E05"
                     /clone_lib="CUGI-OSJNBa"
ORIGIN
```

```
Query Match          6.8%;   Score 182.8;  DB 2;  Length 123269;
Best Local Similarity  61.1%;  Pred. No. 6.9e-34;
Matches 364;  Conservative  0;  Mismatches 227;  Indels  5;  Gaps  4;

Qy    533 ATAAATCACCAAAACTCATGGAGCTTGCTAGTTATAAACTCTAATTCTAAGTTTGGT-GT 591
          ||| |||||||||||||||||  || ||||||  ||||  |||||||||||||||| ||
Db 112780 ATATTCACCAACACTAGTGAAATAATAATTAGTAGTAAACCTTAAACCTTGGATAAC 112839

Qy    592 TTATTGAGTGGATTTGTGTGAAAGTTGGTGTGGTTAGAAAATAGGAGTTAAGGACCGCCAA 651
          |||||||||||| |||||||||  |    ||||| ||| ||||||||||||||||||||||
Db 112840 ATATTTATTTCATTATATGCAAGTCAAGTATTGACGGTCAAATATTAAGGACCATCA 112899

Qy    652 CAAGATCCCCACACTTAGCCCTTTGCTCATCCTCCGAGTAAAGTTCAAGGACTAAGGTGG 711
          |||||||||||| || || |||||| ||||||  |||| | ||||||||||||| |||
Db 112900 ACAGCACCCCCACACTTAATCTTTGCTCGTCCCGAATGAAGGAGGCTATAGACTATACAATAG 112959

Qy    712 AACATCTCCTCAAATGGTACGATGCCTGCGATATAAGTTATTCCAAGCCTCACCTATACAT 771
          ||||| ||||||||||||  || |||||   |||||||||| ||||||||||||||||||
Db 112960 GGTTATGACTTCCTTACTTACTAACATGCAAACAAATTATTCAAAGCCATACCTGTACTT 113019

Qy    772 GT-GAACTTTGAAGTGTCTACCACGGCATCTTGAGCAATGGAAATGGGAACAGATCAGA 830
          || |||| ||||| || || |||||||||| |||||||||| |||| |||||||||||
Db 113020 ATGGAACTTTGTAGTGTCTACCATGGCATCTTGAGCAATCTTGAGCAATTGAGAAATGGGAACAATCTTGA 113079

Qy    831 ATCCAGTCATCTTTACCTCCTCTTGCTTAGATAACTTGGGTTTTGTAAGGTTTCAAATT 890
          ||||||||||||  |||  ||||||||  ||||||  ||||||||||||||||||| ||
Db 113080 CTCAAGTCACCCTAACATTTCATGAAGAACAACTTGGAGTTTTTTTTGTAACACAAAAT 113139

Qy    891 TAAAACATAGTCTTGCTCCTCCCAAATGATTCTTCTCATATAGCTCAATGTGTATGGTTTCTC 950
          || || || ||||| || |  ||||||||| |   |||||||  |||| ||| |||| ||||
Db 113140 TATAACTTTGCTTTGCTCCTCCCAATTCCTTA-ATCACTCAAAGTGTATGATTGATTCCTC 113198

Qy    951 ACCAAGGC--AATGTTTTGCCTCTTTTCATCCTACTTCTCTAATATTTCTTTTGTGGAGCTT 1008
          |||||||   ||||||||||| ||||| |  ||||||||| || || ||||||||||||
Db 113199 ACAAGGCATAGATGTGCCTTTTCTTTTTCCTTACCATCTAATAAGGCTATGTGAAGCTC 113258

Qy   1009 AGGGTAGGGAATGAAAAGGAAGCATACTTGCATTGCATTGTTACTAAGTCAAAACCAA 1068
          ||||| |||| ||||||||  || || ||||||||||||  ||||| ||||||  |||
Db 113259 AGGGCATGGAAAAGAAGTAGAGCATTACTTGCCATCACATATTGCAAAGTCAAATAATTG 113318

Qy   1069 ATCTGAGGAGGAAGCAAGTCATCACAATCTGATCAAGATGTGTGGATATG 1124
```

```
Db      113319 ATCCACAGAATACAAGACATACTTTCTGGTCAAGAAGTGCATGTGTAATGTGTATG 113374

RESULT 7
AP006062
LOCUS       AP006062             176627 bp    DNA     linear   PLN 28-JUL-2004
DEFINITION  Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 9,
            PAC clone:P0415D04.
ACCESSION   AP006062
VERSION     AP006062.2  GI:40363797
KEYWORDS    .
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T., Hattori,M., Sakaki,Y. and Katayose,Y.
  TITLE     Oryza sativa nipponbare(GA) genomic DNA, chromosome 9, PAC
            clone:P0415D04
  JOURNAL   Published Only in Database (2002)
REFERENCE   2  (bases 1 to 176627)
  AUTHORS   Sasaki,T., Matsumoto,T., Hattori,M., Sakaki,Y. and Katayose,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-NOV-2002) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     On Dec 26, 2003 this sequence version replaced gi:25956258.
            Genes were predicted from the integrated results of the following:
            GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH
            (http://www.softberry.com/), GeneMark.hmm
            (http://opal.biology.gatech.edu/GeneMark/), GlimmerM
            (http://www.tigr.org/tdb/glimmerm/glmr_form.html), RiceHMM
            (http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor
            (http://bioinformatics.iastate.edu/cgi-bin/sp.cgi), sim4
            (http://globin.cse.psu.edu/html/docs/sim4.html), gap2
            (http://www.tigr.org/software/glimmerm/), BLASTN and BLASTX. The
            genomic sequence was searched against NCBI NonRedundant Protein
            database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA
            sequence database at RGP or DDBJ. Protein homologies of the coding
            regions were searched against NCBI NonRedundant Protein database
            with BLASTP. ESTs represent the identified cDNA sequences using
            BLASTN with the corresponding DDBJ accession no. and RGP clone ID.
            Full-length cDNAs represent the identified cDNA sequences using
            BLASTN with the corresponding DDBJ accession no.
            A gene with identity or significant homology to a protein is
            classified based on the protein name to indicate the homology level
            such as same name, 'putative' and '-like protein'. A gene without
            significant homology to any protein but with full-length cDNA or
            EST homology (covering almost the entire length of partial
            sequence) is classified as an 'unknown' protein. A gene predicted
            by two or more gene prediction programs is classified as a
            'hypothetical' protein according to IRGSP standard. A gene
            predicted by a single gene prediction program is also classified as
            a probable 'hypothetical' protein and is included as a
            miscellaneous feature of the sequence.
            The orientation of the sequence is from T7 to SP6 of the PAC clone.
            This sequence of P0415D04 clone has an overlap with OSJNBa0018I03
            (DDBJ: AP005746) clone at 5' end with an overlap with OJ1227 D07
            (DDBJ: AP005563) clone at 3' end. Detailed information on overlap
            and assembly quality together with annotation of this entry is
            available at
            http://rgp.dna.affrc.go.jp/GenomeSeq.html.
FEATURES             Location/Qualifiers
     source          1..176627
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="9"
```

```
     gene            /clone="P0415D04"
                     419..1964
                     /gene="P0415D04.1"
     mRNA            join(<419..493,1764..>1964)
                     /gene="P0415D04.1"
                     /note="start and end point are not identified"
     CDS             join(419..493,1764..1964)
                     /gene="P0415D04.1"
                     /note="predicted by GENSCAN etc."
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="BAD20084.1"
                     /db_xref="GI:47496974"
                     /translation="MARAPRPSGTWVAPHVDKHLPCSCKAWRGSTGAALGHGGVELGR
                     RQVAVVGGSGVAAAGGGVVPGVGSGVPAVGGGVVALGLAVVGSESG"
     gene            2957..3285
                     /gene="P0415D04.2"
     mRNA            join(<2957..3090,3213..>3285)
                     /gene="P0415D04.2"
                     /note="start and end point are not identified"
     CDS             join(2957..3090,3213..3285)
                     /gene="P0415D04.2"
                     /note="predicted by GeneMark.hmm etc."
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="BAD20085.1"
                     /db_xref="GI:47496975"
                     /translation="MWEVAELDYRTCQLDLDVSDGCIYSSHGHITSWTGAQQQSHWVS
                     RCLSKLDFPGNTKRGRKQAMNVDGL"
     gene            7914..8360
                     /gene="P0415D04.3"
     mRNA            join(<7914..8025,8119..>8360)
                     /gene="P0415D04.3"
                     /note="start and end point are not identified"
     CDS             join(7914..8025,8119..8360)
                     /gene="P0415D04.3"
                     /note="predicted by GeneMark.hmm etc."
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="BAD20086.1"
                     /db_xref="GI:47496976"
                     /translation="MLAFLVIKILYFRVIAYGIESGKAVNSAAPNKTVFLCAQHPVTNQ
                     GVALMGDDKKLIDGTSKEGVVIQWVIREVGGGSRYPVLTKTNYSNWELLMKVKLKTRA
                     LWHAIEDDVLIHKMR"
     gene            9558..10867
                     /gene="P0415D04.4"
     misc_feature    9558..10867
                     /gene="P0415D04.4"
                     /note="probably inactive due to including stop codon(s) in
                     CDS
                     pseudogene, polyprotein"
                     /pseudo
     gene            complement(11477..13820)
                     /gene="P0415D04.5"
     mRNA            complement(join(<11477..11543,13250..13395,13446..>13820))
                     /gene="P0415D04.5"
                     /note="start and end point are not identified"
     CDS             complement(join(11477..11543,13250..13395,13446..13820))
                     /gene="P0415D04.5"
                     /codon_start=1
                     /product="zinc knuckle domain-like"
                     /protein_id="BAD20087.1"
                     /db_xref="GI:47496977"
                     /translation="MAADLPRRPFPGHPPSPRLYLNPPRASLLRFRHSPLFLPVGI
                     ELAPSSLSAVAAELRSAAVVASDLLPLRHRLQHRRVLADLIHPFVSPADRRNAAVL
                     VDPSHAAASLRFGRHSRRRRRPGRRSAVPVAAGGDHRGAGALPPAGRAGLAALGRAEW
                     LPRGARPSAARALGKRSFGEEVLGGVLAYTPVERL"
     gene            complement(16991..18766)
                     /gene="P0415D04.6"
     misc_feature    complement(16991..18766)
                     /gene="P0415D04.6"
                     /note="probably inactive due to including stop codon(s) in
                     CDS
```

```
gene            pseudogene, polyprotein"
                /pseudo
                complement(join(18986..19067,19552..19651,20082..20217,
                20446..20502))
                /gene="P0415D04.7"
misc_feature    complement(join(18986..19067,19552..19651,20082..20217,
                20446..20502))
                /gene="P0415D04.7"
                /note="hypothetical ORF
                predicted by GENSCAN
                this category is not included in IRGSP standard"
gene            20920..21852
                /gene="P0415D04.8"
misc_feature    20920..21852
                /gene="P0415D04.8"
                /note="polyprotein-like"
gene            22382..26632
                /gene="P0415D04.9"
misc_feature    22382..26632
                /gene="P0415D04.9"
                /note="probably inactive due to including stop codon(s) in
                CDS
                pseudogene, gag-pol polyprotein"
                /pseudo
gene            complement(28342..29877)
                /gene="P0415D04.10"
misc_feature    complement(28342..29877)
                /gene="P0415D04.10"
                /note="probably inactive due to including stop codon(s) in
                CDS
                pseudogene, polyprotein"
                /pseudo
gene            30361..31629
                /gene="P0415D04.11"
misc_feature    30361..31629
                /gene="P0415D04.11"
                /note="probably inactive due to including stop codon(s) in
                CDS
                pseudogene, polyprotein"
                /pseudo
gene            complement(31644..32369)
                /gene="P0415D04.12"
misc_feature    complement(31644..32369)
                /gene="P0415D04.12"
                /note="hypothetical ORF
                predicted by GENSCAN
                this category is not included in IRGSP standard"
gene            complement(join(34597..35101,35470..35510))
                /gene="P0415D04.13"
misc_feature    complement(join(34597..35101,35470..35510))
                /gene="P0415D04.13"
                /note="hypothetical ORF
                predicted by GENSCAN
                this category is not included in IRGSP standard"
gene            join(35864..35918,36014..36564)
                /gene="P0415D04.14"
misc_feature    join(35864..35918,36014..36564)
                /gene="P0415D04.14"
                /note="pseudogene, hypothetical protein similar to putative
                retroelements"
                /pseudo
```

```
Qy      652  CAAGATCCCCACACTTAGCCCTTTGCTCATCCTCGAGTAAAGTTCAAGGACTAAGGTGG  711
                ||  ||||||||||  ||    ||     |||||||||||||||||||||||||||||||
Db    67945  ACAGCACCCCACACTTAATCTTTGCTCGTCCGAATGAAGTGAGGCTATAGACTATACAATAG .68004

Qy      712  AACATCTCCTCAAATGGTACGATGCTGCATAAAGTATTCCAAGCCTCACCTATACAT  771
                |||    |||||||||||||  ||||||||||||||||||||||||||
Db    68005  GGTTATGACTTCCTTACTAATAATCATGCAAACAAATTATTCAAAGCCATTACCTGTACTT  68064

Qy      772  GT-GAACTTTGAAGTGTCTACCACCAGGCCATCTGGGTGGTTGAGAAATGGAACAGATCAGA  830
                || |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    68065  ATGGATCTTTGTAGTGTCACCATGCGATCTTGAGCAATTGAGAAATGGAACAATCTTGA  68124

Qy      831  ATCCAGTCATCTTACCTCCTTGCTTAGATAACTTGGGTTTTGTAAGGTTTCAAATT  890
                ||||||||   ||||||||||||||||||||||     |||||||  ||||||||
Db    68125  CTCAAGTCACCCTAACATTCATGAAGAACAACTTGGAGTTTTTTTTGTAACACAAAT  68184

Qy      891  TAAAACATAGTCTTGCTCCTCAAATGATTCTCATATAGCTCAATGTGTATGGTTTCTC  950
                |||  ||||||  |||||||||||   |||||||||||   ||||||||||||||||
Db    68185  TATAACTTTGCTTTGCTCTCCATTGATTCTCTTA-ATCACTCAAAGTGTATGATTCCTC  68243

Qy      951  ACCAAGGC--AATGTTTTGCTCCTTTCACCCTACTTCTATATATTCTTTGTGGAGCTT  1008
                |||||||   ||||||||||||||||||||||||||||||    ||||||||||||
Db    68244  ACAAAGGCATATGATTGTGCCTTTTCTTTTCCTACCATCTAATAACGCTATGTGAAGCTC  68303

Qy     1009  AGGGTAGGGAATGAAAAGGAAGCATACTTGCATTGCATTGATATGTTACTAAGTCAAAAACCAA  1068
                ||||||||||||    ||||   |||||||||||||||  |||  ||||||||||
Db    68304  AGGGCATGGAAAAGAGTAGAGCATTACTTGCATCACATATTGCAAAGTCAAATAATTG  68363

Qy     1069  ATCTTGAGGAGGAAGCAAGTCATACAATCTGATCAACATGTGCAAGTGTGTGATATG  1124
                |||||||||||   |||||||||||||||||||   ||||||||||||||||||||
Db    68364  ATCCCACAGAAATACAGACATACTTTCTCGGTCAAGGAGTGCATGTGTATGTGTATG  68419
```

```
RESULT 8
AP002970/c
LOCUS        AP002970                139201 bp    DNA        linear    PLN 31-AUG-2004
DEFINITION   Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1,
             PAC clone:P0504D03.
ACCESSION    AP002970
VERSION      AP002970.2  GI:1161359
KEYWORDS     BA000010.
SOURCE       Oryza sativa (japonica cultivar-group)
ORGANISM     Oryza sativa (japonica cultivar-group)
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
             Ehrhartoideae; Oryzeae; Oryza.
REFERENCE    1
AUTHORS      Sasaki,T., Matsumoto,T., Yamamoto,K., Sakata,K., Baba,T.,
             Katayose,Y., Wu,J., Niimura,Y., Cheng,Z., Nagamura,Y.,
             Antonio,B.A., Kanamori,H., Hosokawa,S., Masukawa,M., Arikawa,K.,
             Chiden,Y., Hayashi,M., Okamoto,M., Ando,T., Aoki,H., Arita,K.,
             Hamada,M., Harada,C., Hijishita,S., Honda,M., Ichikawa,Y.,
             Idonuma,A., Iijima,M., Ikeda,M., Ikeno,M., Ito,S., Ito,T., Ito,Y.,
             Ito,Y., Iwabuchi,A., Kamiya,K., Karasawa,W., Katagiri,S.,
             Kikuta,A., Kobayashi,N., Kono,I., Machita,K., Maehara,T.,
             Mizuno,H., Mizubayashi,T., Mukai,Y., Nagasaki,H., Nakashima,M.,
             Nakama,Y., Nakamichi,Y., Nakamura,M., Namiki,N., Negishi,M.,
             Ohta,I., Ono,N., Saji,S., Sakai,K., Shibata,M., Shimokawa,T.,
             Shomura,A., Song,J., Takazaki,Y., Terasawa,K., Tsuji,K., Waki,K.,
             Yamagata,H., Yamane,H., Yoshiki,S., Yoshihara,R., Yukawa,K.,
             Zhong,H., Iwama,H., Endo,T., Ito,H., Hahn,J.H., Kim,H.I., Eun,M.Y.,
             Yano,M., Jiang,J. and Gojobori,T.
TITLE        The genome sequence and structure of rice chromosome 1
JOURNAL      Nature 420 (6913), 312-316 (2002)
PUBMED       12237376
MEDLINE      12447438
REFERENCE    2  (bases 1 to 139201)
AUTHORS      Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE        Direct Submission
JOURNAL      Submitted (29-NOV-2000) Takuji Sasaki, National Institute of
             Agrobiological Sciences, Rice Genome Research Program; Kannondai
             2-1-2, Tsukuba, Ibaraki 305-8602, Japan
             (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
```

```
Query Match     6.8%;   Score 182.8;  DB 8;  Length 176627;
Best Local Similarity    61.1%;  Pred. No. 7.3e-34;
Matches 364;  Conservative    0;  Mismatches 227;  Indels    5;  Gaps    4;

Qy     533  ATAAATCACCAAAACTCATGGAGCTTGCTAGTTATAAAACTCTAATTCTAAGTTTGGT-GT  591
               ||| |||||||| ||              ||||  ||||||||||||||||
Db   67825  ATATCTCACCAACACTAGTGGAAATAATTAGTAGTAAAACCTTACCACTAAGTTGGATAAC  67884

Qy     592  TTATTTGAGTGGATTTGTGTGAAAGTTGGTGGTTAGAAATAGGAGTTAAGGACCGCCAA  651
               |||| |||||||||||  |||||||||||||||||  |||||||||||
Db   67885  ATATTTTATTTCATTTATATGCAAGTATTGACGAGTTGACGGTCAAATATTATGATTTAAGGACCATCA  67944
```

```
COMMENT     Tel:81-298-38-7441, Fax:81-298-38-7468)
            On Feb 28, 2001 this sequence version replaced gi:11526597.
            Genes were predicted from the integrated results of the following:
            GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor
            (October 1998 version). The genomic sequence was searched against
            NCBI NonRedundant Protein database, nr
            (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
            RGP. Protein homologies of the coding regions were searched against
            NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent
            the identified cDNA sequences using BLASTN 2.0 with the
            corresponding DDBJ accession no. and RGP clone ID.
            A gene with identity or significant homology to a protein is
            classified based on the protein name to indicate the homology level
            such as same name, 'putative-' and '-like protein'. A gene without
            significant homology to any protein but with EST homology (covering
            almost the entire length of partial sequence) is classified as an
            'unknown' protein. A gene predicted with a gene prediction program
            is classified as a 'hypothetical' protein.
            The orientation of the sequence is from SP6 to T7 of the PAC clone.
            Detailed information on overlap and assembly quality together with
            annotation of this entry is available at
            http://rgp.dna.affrc.go.jp/GenomeSeq.html.
FEATURES             Location/Qualifiers
     source          1..139201
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="1"
                     /clone="P0504D03"
     gene            join(2371..2591,2628..2925)
                     /gene="P0504D03.1"
     CDS             join(2371..2591,2628..2925)
                     /gene="P0504D03.1"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAB32955.1"
                     /db_xref="GI:13161360"
                     /translation="MAATVAVARLMAARREDIGEAKDDAVGDGEADAAAVDRTARHVL
                     GIRSWTTVAEFTPSADLTTTVEPAKGTRGWRSARQWRAMAGDGEEEDATAVGEEDNA
                     EAVWRRQCSSGECBGGRRGGDGBEAMQWRWVGRRRTRRRWQVREIRRGGVAGRGSRA
                     VVASTAGEVFVR"
     gene            complement(join(3966..4098,5324..5442,5530..5649,
                     5655..5711))
                     /gene="P0504D03.2"
     CDS             complement(join(3966..4098,5324..5442,5530..5649,
                     5655..5711))
                     /gene="P0504D03.2"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAB32956.1"
                     /db_xref="GI:13161361"
                     /translation="MQEQLSSLLDIELVDAPPHSDASLSPHPHRRSRFAPVLIDVE
                     GVIPGFMVREMIRSRPTRRRCTQPRAILAEMRTHTNCGRWTGARRAVLALVNGGRLNQ
                     HVGPIIICILIRIALIPTVHCDSPSHSNSMETTLHTSDG"
     gene            join(14660..14794,15810..15929,16670..16752,16842..16953,
                     17025..17129,17244..17294,20044..20133,20217..20325,
                     20400..20548,20646..20699,22252..22455)
                     /gene="P0504D03.3"
     CDS             join(14660..14794,15810..15929,16670..16752,16842..16953,
                     17025..17129,17244..17294,20044..20133,20217..20325,
                     20400..20548,20646..20699,22252..22455)
                     /gene="P0504D03.3"
                     /codon_start=1
                     /product="putative ARG1 protein (Altered Response to
                     Gravity)"
                     /protein_id="BAB32957.1"
                     /db_xref="GI:13161362"
                     /translation="MASSGKMEGPSAPAMRRDPYEVLSVPRDSSDQEIKSAYRKLALK
                     FHPDKNASNPEASELFKEVAYSYSILSDPEKRRQYDTAGFEALENEGMDMEIDLSNLG
                     TVNTMFAALFSKLGVPIKTTVSPNVLEEAMSGTVTVRPLPVGSSATGKVDKQSAHFYG
                     VTISEEQAQSGIVRVTSAAQSKFLLFFEQRINGGYGLALQEDSQKTGKVTSAGMYF
                     LHFQVYRMDSTVNALAMAKDPEAAFFKRLEGLQPCEVSALKSGTHIFAVYGDNFFKPA
                     SYTIEAMCAKSYEDTTQRLKEIESKILEKRNDLRQFETEYRKALARFQEVTNRYTQEK
                     EAVDDMLRERDDIHSSFTTERTMVNSVGAGSSSRYPTESPENGNIDGKDKSSKKKWF
                     NLNLNRSDKKA"
     gene            24163..27000
                     /gene="P0504D03.4"
     CDS             24163..27000
                     /gene="P0504D03.4"
                     /note="contains ESTs AU101970 (S3382),D41106 (S3382)"
                     /codon_start=1
                     /product="putative adaptin"
                     /protein_id="BAB32958.1"
                     /db_xref="GI:13161363"
                     /translation="MASAPPAAPAPAPSLVDTLFQRSLDDLVKSLRADPSAAGESAAV
                     ARALSEIHREIRAPDTATKAVALQKLTYLSSLHFAPVASHPLAFPAIELIASPQLPHR
                     RLAYLAASLSLHPASLSLLPLAITHQLHKDLSPTSSAANHVTALALQLLASPAAAA
                     PDLPVHLAHDLVPHLSRGSPRAIAAARVIAASPSAAVPVLFKPLAACLASPDPRASA
                     AAAAFCELSAPPADAAPLPLAPDLYNLLTTSRSNWALIKVLKVFARLAPLESRLAA
                     RIVDPVCQLLARSAAMSLTFECIRTVLTALPAHDAAVRLAIGKAKEFLAAEDDPNLRY
                     LGLLALGMLGPAYASTLHECRGVIAQSLGDADSNICREALHLMMGLIDDSNVTDIAGM
                     LVSHASKSDPEFANDILGAVLSACGRNVYELVSDFDWYSLLTDMGRNLHCAQGDEIG
                     RQLVDVGLRVQDARPELVHSSRTLLIDPALLGNHLLCPVLSAAAWVSGEYVNFTKDPV
                     ELVEALLQPRISLLPMSVRAVYIQAVLKVITTGPSLKSVHMINLIETTVGPLVECNEVEVLE
                     QTVSGGSDAPIGSSNGQITVPRMMEKDPFSLKSVVHMINLIETTVGPLVECNEVEVLE
                     RARNLIGFVYSLRBIQELKESKFDDDKHSRVKEIVKNMQTVLSHEIGPVSLNAQEKVS
                     LPDDLVLNENLAELVDIISEDDTTLSSSIVFYPRSCGSVETRDEPALSLGSSLLSEH
                     RKRHGLYYLPTGKAEDGVPDYPHANDPLLPASSESALDDKLKTIQPVTCGKKPKAVKS
                     RPKVVKLDGEDFLSSMVASASVPKEDSLSGAVRGVLLGRDLKPSSSQKASDKAYEGII
                     NKMDSGESSGWKNNVDADFVGHTSSSRPSIQOSHDKESTNPLESDGKEARKHRSR
                     SGHRQGKHKHRERHCTQPDVPQAPIIQDFLL"
     gene            complement(join(27845..27959,28206..28340,29236..29303))
                     /gene="P0504D03.5"
     CDS             complement(join(27845..27959,28206..28340,29236..29303))
                     /gene="P0504D03.5"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAB32959.1"
                     /db_xref="GI:13161364"
                     /translation="MWPDLTEAATSGGIGKMRNPGRSSLTRGKEKEQIGPLVHPSNRQ
                     FIGLTAGAIIYLLHVRQNRTSCERLQFGGPNCERDTGSHLVVDVASSGQMTNHRTVKA
                     LRA"
     gene            complement(join(34996..36563,37713..37869))
                     /gene="P0504D03.6"
     CDS             complement(join(34996..36563,37713..37869))
                     /gene="P0504D03.6"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAB32960.1"
                     /db_xref="GI:13161365"
                     /translation="MLDALWKATLRTLRKSAVAPWASADAVGPAWGVPVLVASGALNA
                     PCEMSMLGALPPEATLRRSTMRRSPPPAAAQPRPPTRAASGEEGDPGKDEVAA
                     PPKGDSALPRSGSGFGSGNAAAPDGPAKSPMRPKKRVTAPPSSDHGSRENLAAGSPSN
                     EVPTPQGDDAAQEDLPESPPKSKVNRKKKKALRAGDSGKVAAPDDPAEPTRPQQDV
                     GEVDGRLAESEQEVPPNAGKAKKASAQARKPSAKOHAAAQEBEDGLMAEABEEVA
                     PWQGDEDRGKGPLPQRKSKRVAALSGSICPPDRKRAKIVDAXPGFGRKWNGNDEIMI
                     LEALVDQIRSGGNVPQEPGHPLFHELVQRLEGRTFNHSDVRDKVRSLKRRYNDVVLSG
                     LAITKDHDLQLHELSCEIWGRSVAHAGDDKQRCLARDEQSSLARDEQKSFARDEEKSL
                     ARDEQSSLARDEQKSFARDEEKSLARDEKSLASDEQRSFDDMCKQFPLLAKEIKVLME
                     GQPAIMELFPRLDGDQVVAIEKKLENLRWIDMKRKKNAVKMAKIRKGLJYKLEGAAI
                     LADGNMIH"
     gene            complement(join(40009..40028,41743..41887))
                     /gene="P0504D03.7"
     CDS             complement(join(40009..40028,41743..41887))
                     /gene="P0504D03.7"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="BAB32961.1"
                     /db_xref="GI:13161366"
                     /translation="MANLPRVMGGFSHGDGVASNGTSPVPTITVGLPSCTDQGIGPEW
                     KSALESDLHR"
     gene            join(43590..43634,43673..43894)
                     /gene="P0504D03.8"
     CDS             join(43590..43634,43673..43894)
                     /gene="P0504D03.8"
```

gene
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB32962.1"
/db_xref="GI:13161367"
/translation="MERERQRGPDVLLLVEMGTEGEGSVRPERVKADGEGDSAGAGGG
GRGEGGGGEVGVGRGRRRGGRCRGGEESKLLSRAWIRPEEKMRG"
50026. .50735
/gene="P0504D03.9"

misc_feature   50026. .50735
/gene="P0504D03.9"
/note="probably inactive due to frameshift(s) in CDS
pseudogene
similar to Oryza sativa Sgt1"
/pseudo

gene   complement(join(51394. .51430,52618. .52850))
/gene="P0504D03.10"

Query Match   6.7%; Score 181.4; DB 8; Length 139201;
Best Local Similarity   63.2%; Pred. No. 1.6e-33;
Matches   367; Conservative   0; Mismatches 196; Indels   18; Gaps   5;

Qy   486 TTTCTTGGCCCATCTTGGTATTTCACAAATGTCCCCTACAAATGATAAATCACCAAA   545
            |||||| |||| ||||| ||| ||||| |||||||| |||| ||||||||||||
Db   72222 TTTCTTCGCATGGTGGGGTGGTTTTTGTTGTCATAAATTCACCTTTATACAAATATTCACCAAG 72163

Qy   546 ACTCATGGAGCTTGCTAGTTATAAACTCTAATTCTCAAGTTGGTGTTTATTTGAG-----   600
            ||| ||| || ||||| ||| | ||| || || ||||| | ||||| ||||
Db   72162 ACTTATGGAAAATGGTTAGTAATAACCACTACCCTCTCAAGTTGATATCTATTTTATATTCT 72103

Qy   601 ----TGGATTTTGTGTGAAAGTGGTGGTGTGTTAGAAATAGGAGTGAAGGACCGCCAACAAG   655
                || ||||||| | ||||||||||||| ||| ||| ||||| || ||||||||||||||
Db   72102 TTCATTGCTTTTATGCATGAATTCACAATCAGATCTGGTACTTAAGGACTGTTAACAAC 72043

Qy   656 ATCCCCACACTTAGCCCTTTGCTCATCCTCGAGTAAAGTTCAAGGA-CTAAGGTGGAAC   714
            ||||||| |||||| ||||| |||||| ||||| |||||| ||||| || || |||
Db   72042 A-CCCCCACACTTAG-CTTTTACTCCGTCACTGAGTAAAGATCATGGATCAACAAAGAT 71985

Qy   715 ATCTCCTCAAAATGTACGATGCCTGCATATAAGTTATTCCAAGCTCACCTATACATGTG   774
            ||||| ||| ||||| || || |||| ||| || |||||||| |||| || ||||
Db   71984 ATGATTCTTTGTAATGATAACCTGCATAATAATTATTCCAAGCCATACTTGTACTTGTG 71925

Qy   775 AACTTGAAGTGTCCACCACCGCCATCTTGGGTGGTTGAGAAATGGAACAGATCAGAATCC   834
            || ||||||| | ||||| |||||||| ||||||| | || |||||| |||| ||||
Db   71924 AATCTTGAAGTATCTACCACCGCCATCTTGGGTGATTGAAGAACGGAACCGTCTTGATTCA 71865

Qy   835 AGTCATCTTTACCTCCTTGCTTAGATAACTTGGGTTTTGTGTAAGGTTTCAAATTTAAA   894
            || | ||||| |||||| ||| |||||| ||||| |||| ||||| ||||| |||
Db   71864 AGTTACCTCCACTCCTCCTGATGAGTAGTACATTGGGTTTTCGTTTGCTCATAAAAACATA 71805

Qy   895 ACATAGTTCTTGCTCTCAAATGATTTCTCTATAGCTCAATGTGTATGGTTTCTCCACCA   954
            | | || |||| |||| |||| ||||| || ||| |||||||| | || |||||||
Db   71804 ACTCTACTTTATTCCTTAATTGTTTGTCTCGGATCACTCAAATGTATTAATTCTTCACCA 71745

Qy   955 AGGCAATGTTTTGCC------TCTTTTCATCCTACTCACTACTTCTTTTGTGGAGCTTA   1009
            ||| |||| || || |||||| |||||| |||| ||| ||||||||||| |||
Db   71744 AGGTATTGATTTTGCGTTTTTTTCTTTATCCTTCTAATAAGGATTATGTCGAGCTCA 71685

Qy   1010 GGGTAGGGAATGAAAAGAAGCATACTTGCATTCGCATATGT   1050
            |||||||| |||| ||||||||||| |||||| | |||
Db   71684 GGGTAGGGGGAAAGAAAAGAACAAAAACATTACTACTACACATATCT 71644

RESULT 9
AP003213/c

COMMENT   On Jun 19, 2001 this sequence version replaced gi:13620982.
            Genes were predicted from the integrated results of the following:
            GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor
            (October 1998 version). The genomic sequence was searched against
            NCBI NonRedundant Protein database, nr
            (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
            RGP. Protein homologies of the coding regions were searched against
            NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent
            the identified cDNA sequences using BLASTN 2.0 with the
            corresponding DDBJ accession no. and RGP clone ID.
            A gene with identity or significant homology to a protein is
            classified based on the protein name to indicate the homology level
            such as same name, 'putative-' and '-like protein'. A gene without
            significant homology to any protein but with EST homology (covering
            almost the entire length of partial sequence) is classified as an
            'unknown' protein. A gene predicted with a gene prediction program
            is classified as a 'hypothetical' protein.
            The orientation of the sequence is from M13rev to -21M13 of the BAC
            clone. This sequence of OSJNBa0054L14 clone has an overlap with
            P0504D03 clone (DDBJ: AP002970) at the position 50,801 to 150,022
            of 3' end. The sequence of this clone ends at the position 99,222
            of P0504D03. Detailed information on overlap and assembly quality
            together with annotation of this entry is available at
            http://rgp.dna.affrc.go.jp/GenomeSeq.html.
FEATURES   Location/Qualifiers
     source   1. .150022
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="1"
     gene   complement(join(14480. .14558,15653. .15771,16436. .16545,
                     16663. .16858))
                     /gene="OSJNBa0054L14"
     CDS   complement(join(14480. .14558,15653. .15771,16436. .16545,
                     16663. .16858))
                     /gene="OSJNBa0054L14.1"
                     /note="contains ESTs
                     AU01311(E10180),AU161063(C53367),C19264(E10180),AU063176
                     (C53367),C19811(E10967)
                     unknown protein"
                     /codon_start=1

gene   complement(join(37198..37348,37669..37798,37941..38118,
        38273..38382,38455..38573,38690..38951,39586..39901,
        40000..40296))
        /gene="OSJNBa0054L14.7"

CDS   complement(join(37198..37348,37669..37798,37941..38118,
        38273..38382,38455..38573,38690..38951,39586..39901,
        40000..40296))
        /gene="OSJNBa0054L14.7"
        /note="contains EST AU096093(S12136)
        unknown protein"
        /codon_start=1
        /protein_id="BAB60914.1"
        /db_xref="GI:14495195"
        /translation="MAATSSSSSPCMIASLRSSHRCRLSPSATATSPPRLRTLPPRRC
        RRNPSSSSSRQAAAISMAPANPRHRLIAPIHAAAAGGGGAGATSPSGLPSLVG
        IAHLLVSLGIVLASDKFLKQAFAAASIKFPSALFGMFCVFSVLVLDAFAPALAKGFM
        DFFEPATLFIQRWLPLFYVPSLVVLPLAVRDVPAASGLKIFLIIFGGWFASLMVAGYT
        ALTVRKIVQTQLIPAEPMSKPSPFATLEFWAWGAVFVASFAVAYVNPTALGTTARTCL
        PFLLASTVLGYWGCLDPSGVRVLMPPGLDYISLGLVDED
        LTKAPSNPGAGDVLMGFLGSVIISFAFSMFKQRKLVKRHAAEIFTSIAIASTFSLYST
        PDMTPFPANRYMATLTPIEGYIEKVRDAAKYSVKEKLR"

gene   complement(join(23041..23226,23965..24064,24160..24356,
        25077..25157,25287..25733,26015..26161))
        /gene="OSJNBa0054L14.3"

CDS   complement(join(23041..23226,23965..24064,24160..24356,
        25077..25157,25287..25733,26015..26161))
        /gene="OSJNBa0054L14.3"
        /note="contains ESTs
        AU030681(E60106),AU030680(E60106),C72766(E2205)
        unknown protein"
        /codon_start=1
        /protein_id="BAB60910.1"
        /db_xref="GI:14495191"
        /translation="MSRRLIFKGLTINPALASGMTCQHHQLQQHAPVSCTAKGKAKLKS
        GQQLKRNTIGAKKGGAPSTGGGGGGRGRREAIERITQIAESCLNASTPLRHLSPXER
        LREAKRELGLISKERQRELDLLAKAKASKGTREGDGGRVLMGPPGLDYISLGLVDED
        AIPKYELTVEDGRRLAKQYSQVLMRRHRARQTAESSLISLKKEAIAALPEKLRAAAMI"

gene   join(27375..27382,27782..28478)
        /gene="OSJNBa0054L14.4"

CDS   join(27375..27382,27782..28478)
        /gene="OSJNBa0054L14.4"
        /note="contains ESTs
        AU108864(C51670),C96925(C50129),C26812(C50129),C27351
        (C51670),D41570(S4148)"
        /codon_start=1
        /product="putative 26S proteasome subunit"
        /protein_id="BAB60911.1"
        /db_xref="GI:14495192"
        /translation="MALQYVEAQRQARPDLADAYADLADLYQRKLWHQLTLKLDHFLQ
        LPAAQTGDTIIQLYNNFISDFETKINLLKLAHFAVIASRQYPDKDAISFLEGVTAKL
        RETRERRIDEPVLYVKMQIAAINLEKGDQKECKLLDEGKSTLDSMTDVDPTVHASFY
        WISSQYHKACQFRAEFYKNALLYLAYTTVESLSESFKLDLAFDLSLAALLGDNIYNFG
        ELLAHPIINSLIGTKVEWVYHMLQAFNTGNLALYQELCKVHNAALSAQPALVQNERKL
        LEKINILCLMEIIFTRPSEDRTPLSVIAERTKLSISDVEYLLMKSLSVHLIEGIIDE
        VDSTVHVSWVQPRVLGIPQVKALRERLDAWVGKVHTTLLSVEAETPDLVAA"

gene   complement(19024..19800)
        /gene="OSJNBa0054L14.2"

CDS   complement(19024..19800)
        /gene="OSJNBa0054L14.2"
        /note="contains ESTs
        AU030681(E60106),AU030680(E60106),C72766(E2205)
        unknown protein"
        /codon_start=1
        /protein_id="BAB60909.1"
        /db_xref="GI:14495190"
        /translation="MASPSAPGPKLQKAMVAVDESEFSHHALEWALRNLAPTIAPPLL
        VLTVQPLLPLGVYSAAASRFGSPLGTPVVAPELIKAMQEQQQLSQALLDKAKQICAQHG
        VAVETMIKVGDPKEMICQAAEESKVDLLIVGSHSRGPVQRLFLFGSVSNYCMHHSKCPV
        LVVKKQE"

gene   join(41004..41132,44330..44615,44748..44833)
        /gene="OSJNBa0054L14.8"

CDS   join(41004..41132,44330..44615,44748..44833)
        /gene="OSJNBa0054L14.8"
        /note="hypothetical protein"
        /codon_start=1
        /protein_id="BAB60915.1"
        /db_xref="GI:14495196"
        /translation="MAKTFTAKTFTDYYYSPNRLIPMLIGFASNKYQAIPQLTIREK
        TGKGREKNRKTKTQNRKRTEKREKQKKTKTEREPKKTTEBKQKPKQKNRKKKTHARA
        PFAASFACDTCPVVQGRARATGRRLAFSKLHPSNPKGQLGEIASLAPYRTTVVRQGTY
        RTTTVR"

gene   join(28891..30563)
        /gene="OSJNBa0054L14.5"

misc_feature   complement(28891..30563)
        /gene="OSJNBa0054L14.5"
        /note="probably inactive due to frameshift(s) in CDS
        probably inactive due to no initiation/termination codons
        in CDS
        pseudogene, gag-pol polyprotein"
        /pseudo

gene   join(45358..45531,45640..45719,46726..46762)
        /gene="OSJNBa0054L14.9"

CDS   join(45358..45531,45640..45719,46726..46762)
        /gene="OSJNBa0054L14.9"
        /note="hypothetical protein"
        /codon_start=1
        /protein_id="BAB60916.1"
        /db_xref="GI:14495197"
        /translation="MTTPPSATRSLLPSPRYTSLRASSLLPVVATMVCPHRRGGSPF
        PLEPCLDYPDGAGVDDASGYIASKPSTAAGILEVMRRPEBDSSADNHTYKSYN"

gene   join(33259..33262,35302..35482,35726..35819)
        /gene="OSJNBa0054L14.6"

CDS   join(33259..33262,35302..35482,35726..35819)
        /gene="OSJNBa0054L14.6"
        /note="hypothetical protein"
        /codon_start=1
        /protein_id="BAB60913.1"
        /db_xref="GI:14495194"
        /translation="MATSGGGDARGREARRRRAAVLRAGREALGRGAASSRARRTGR
        RARRGVRTRREARQPRGQLTGKVAVITGASGIGRATAEEFIVPNNVF"

gene   complement(join(33198..33455..38573,38690..38951,39586..39901,
        40000..40296))
        /gene="OSJNBa0054L14.7"

CDS   complement(join(33198..33455..38573,38690..38951,39586..39901,
        40000..40296))
        /gene="OSJNBa0054L14.7"

        /translation="MQYAHGEDELGDGRSVWAVGEVRRAWRGGGHIVDSIGEVRTAA
        HGRRSQLWVAERARGRRRPPEAEANRVHGDARWPVGSPSSLWAHGDARGHHRHCSSSS
        EPRRRAPSHLPVPGSAMETGGRSTEDAGVNGDRESRSTTPDPNASSVMTHGAWVGGR
        ACVGSGSGGHDSNNAGTLVPAPVPASSSVSAGRPSRRPRWWRMWLRESEAVVMESTE
        QSWSCDDAFSRGPIFW"

Query Match   6.7%; Score 181.4; DB 8; Length 150022;
Best Local Similarity   63.2%; Pred. No. 1.6e-33;
Matches 367; Conservative   0; Mismatches 196; Indels 18; Gaps   5;

Qy   486 TTTCTTGGCCCATCTTGGTATTTCACAAATGTCCCCCTACAAATGATAAAATCACCAAA 545
              ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   123022 TTTCTTGCGATGGTGGGTGTTTTTGTTCATAATCGTCATTTATACAAATATTCACCAAG 122963

Qy   546 ACTCATGGAGCTTGCTAGTTATAAACTCTATTCTAAGTTGGTGTTTATTTGAG------ 600
              || ||||||| |||||| |||| |||| || |||||||| |||||||||||
Db   122962 ACTTATGGAAAATGGTTAGTAATAACCACTACCCTCTAAGTTTGATATCTATTTTATATTCT 122903

Qy   601 -----TGGATTTGTGTGAAAGTTGGTGGTTAGAAATAGGAGTTAAGGACCGCCAACAAG 655
                   ||| ||| ||||||||| ||||||| ||||||||||| ||||||||||||||
Db   122902 TTCATTTGCTTTTATGCATGAATTCACAATCAGATCTGGTTACTTAAGGACTGTTAACAAC 122843

Qy   656 ATCCCCCACACTTAGCCCTTTGCTCCATCCTCGAGTAAGTTCAAGGA-CTAAGGTCGAAC 714
              ||||||||||||||||| || |||||||||| || ||||||||||| |||||||||||
Db   122842 A-CCCCCACACTTAG-CTTTTACCTCACTGAGTAAAGATCATGGATCACAACAAAGAT 122785

Qy   715 ATCTCCTCAAATGGTACGATGCCTGCATATAAGTTATTCCAAGCCTCACCTATACATGTG 774
              ||| || ||||||| || ||| |||| |||| ||||||||| |||| | ||
Db   122784 ATGATTCTCTTTGTAATGATAACCTGCATAATAATTATTCCAAGCCCATACTTGTACTTGTG 122725

Qy   775 AACTTTGAAGTGTCTCACCACGCCATCTTGGGTGGTTGGAAATGGAACAGATCAGAATCC 834
              |||||||||| |||||||| ||||||| ||| |||||||||||||||| |||
Db   122724 AATCTTGAAGTATCCTCACCGTCATCTTTGGGTGATTGAAGAGACGGACGGGTCTTGATTCA 122665

```
Qy    835 AGTCATCTTTACCTTCTTGCTTAGATAACTTGGGTTTTTGTAAGGTTTTCAAATTTAAA 894
          ||| ||| ||| |||||| |||||| |||||||||||||||||||| |||||
Db 122664 AGTTACCTCCACTCCTCCTGATGAGACATTGGGTTTTTCGTTTTGTTCCATAAAAACATA 122605

Qy    895 ACATAGTCTGCTCCTCCAAATGATTTCTCTCATATAGCTCCAATGTGTATGGTTCTCACCA 954
          ||| |||                 ||||||  |||||| |||||
Db 122604 ACTTCACTTTATTCCTTAATTGTTTGTCTCGGATCACTCCAATGTGTATAATTCTTCACCA 122545

Qy    955 AGGCAATGTTTTGCC-----TCTTTCATCCTCACTTCTAATATTTCTTTTGTGGAGCTTA 1009
          ||| |||         ||| |||||||||||||||| |||||   ||||||||
Db 122544 AGGTATTGATTTTGCGTTTTTTTTCTTTATCCTCACTTCTTAATAAGGATTATGTGGAGCTCA 122485

Qy   1010 GGGTAGGGAATGAAAAGGAAGCATACTTGCATTGCATATGT 1050
          |||||||                       ||||||||||
Db 122484 GGGTAGGGGAAAGAACAAAAACATTACTTACATCACACATATCT 122444
```

RESULT 10
AC134345
LOCUS   AC134345   136858 bp   DNA   linear   PLN 08-MAY-2004
DEFINITION Oryza sativa (jpaonica cultivar-group) chromosome 5 PAC clone
            P0012A07, complete sequence.
ACCESSION   AC134345
VERSION   AC134345.2 GI:47084429
KEYWORDS   HTG.
SOURCE   Oryza sativa (japonica cultivar-group)
  ORGANISM Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1 (bases 1 to 136858)
  AUTHORS   Chow,T.-Y., Hsing,Y.-I.C., Chen,C.-S., Chen,H.-H., Liu,S.-M.,
            Chao,Y.-T., Chang,S.-J., Chen,H.-C., Chen,S.-K., Chen,T.-R.,
            Chen,Y.-L., Cheng,C.-H., Chung,C.-I., Han,S.-Y., Hsiao,S.-H.,
            Hsiung,J.-N., Hsu,C.-H, Huang,J.-J., Kau,P.-I., Lee,M.-C.,
            Leu,H.-L., Li,Y.-F., Lin,S.-J., Lin,Y.-C., Wu,S.-W., Yu,C.-Y.,
            Yu,S.-W., Wu,H.-P., Shaw,J.-F., *McCombie,W.R., *Zutavern,T.,
            *Spiegel,L., *de la Bastide,M., *Muller,S., *Nascimento,L.,
            *Balija,V., *Bell,M., *Miller,B., *Katzenberger,F., *Andrade,M.V.,
            *Dike,S., *O'Shaughnessy,A. and *Palmer,L.
  TITLE   Oryza sativa (japonica cultivar-group) chromosome 5 PAC clone
            P0012A07, complete sequence
  JOURNAL   Unpublished
  REMARK   *Lita Annenberg Hazen Genome Center, Cold Spring Harbor Laboratory,
            1 Bungtown Road, Cold Spring Harbor, NY, 11724, USA
REFERENCE   2 (bases 1 to 136858)
  AUTHORS   Chow,T.-Y. and Hsing,Y.-I.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (26-SEP-2002) Institute of Botany, Academia Sinica, 128,
            Section 2, Academia Road, Nankang, Taipei 11529, Taiwan
REFERENCE   3 (bases 1 to 136858)
  AUTHORS   Chow,T.-Y. and Hsing,Y.-I.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (08-MAY-2004) Academia Sinica, Institute of Botany, 128,
            Section 2, Academia Road, Nankang, Taipei 11529, Taiwan
COMMENT   On May 8, 2004 this sequence version replaced gi:22321366.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. This PAC clone was
            sequenced to phase II by the Academia Sinica and was completed to
            phase III by the Cold Spring Harbor Laboratory Genome Center.
            Clone P0012A07 overlaps clone OSJNBa0030J05 (AC137746) from base 1
            to base 18791. The overlap is from base 128562 to base 147352 on
            OSJNBa0030J05.
FEATURES   Location/Qualifiers
     source   1..136858
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="5"
                     /clone="P0012A07"
                     /clone_lib="MboI"
     misc_feature   52890..53065
                     /note="The assembly is covered by high quality sequences
                     derived from a PCR product."
     misc_feature   78310..78393
                     /note="The sequence is a dinucleotide (TA) repeat in which
                     the exact number of TA pairs is unknown."
     misc_feature   81730..81985
                     /note="The assembly is covered by high quality sequences
                     derived from a PCR product."
     misc_feature   126085..126089
                     /note="An E. coli insertion element was excised from the
                     assembly at this location."

ORIGIN

Query Match   6.5%; Score 175.8; DB 8; Length 136858;
Best Local Similarity 62.1%; Pred. No. 3.9e-32;
Matches 378;   Conservative   0; Mismatches 222; Indels 9; Gaps   6;
```

```
Qy    529 AATGATAAATCACCAAAACTCATGGAGCTTGCTAGTTATAAAACTCTAATTCTAAGTTGG 588
          ||| ||||||  |||||| ||| ||| |||  |||| ||||| |||  |||||||
Db 113515 ACTCATATTTCACCAACACTAGTGAAATGATTAGTAATAAAATCTTACCACTAAGTTGGA 113574

Qy    589 T-GTTTATTTGAGTGGGATTTTGTGTGAAAGTGGTGGTTAGAAATAGGAGTAAGGACCG 647
          | ||||||||  |||||||||| |||  |||   |||| ||| ||||| ||||||
Db 113575 TACCATATTTTATTTCATTATATGCAGGTATTGATGATTGTCAAATATTATTATTAAGGACC 113634

Qy    648 CCAACAAGATCCCCACACTTAGCCCTTTGCTCATCCTCGAGTAAAGTTCAAGGACTAAG 707
          ||| |||  |||  ||||||||||||  |||||  |||| ||||| |||||||
Db 113635 GTCAACAACACCCCCACACTTAACCTTTGCCCGTCCGAGTGAAGGCTATGGACTATACA 113694

Qy    708 GTGGAACATCTCCTCAAATGGTACGAATCGCATATAAGTTATTCCAAGCCTCACCTAT 767
          ||  ||||| |||    |||  || |||||| ||| ||| ||||| ||||||||
Db 113695 ATAGAGTTTTGTCTTCCCTACGAATAATCATGCAAACAAATTATTTAAAGTCATACCTGT 113754

Qy    768 ACATGT-GAACTTTGAAGTGTCTACCACGCCATCTTGGGTGGTTGGAGAAATGGAACAGAT 826
          || ||| ||||||| |||||| |||||||||||||  |||||||||||||||||||
Db 113755 ACTTGTGGATCTTTGTAGTGTCTACCACCATGGAGTCCTGGGCAATTGAGAATGAGAATAATC 113814

Qy    827 CAGAATCCAGTCATCTTTACCTCTCTTGCTTAGATAACTTGGGTTTTTGTAAGGTTTTCA 886
          |||||  |||||| ||||||||    ||| ||| |||  ||||||||||||  ||||
Db 113815 TTGACTCAAGTCACCCTAACTTTTCATGAAGA-ACAACTTGGAGTTTTTCACGTAACACA 113873

Qy    887 AATTAAAACATAGTCTTGCTCCTCCAAATGATTCTCTCATATAGCTCCAATGTGTATGGTT 946
          ||||||  |||||  |||| || |||| ||||||||  ||||| ||||  |||||||
Db 113874 AAATTATAACTTTGCTTTGCTCCTCTCG-ATCACTCCAAAGTGTATGATT 113932

Qy    947 TCTCACCAAGGCAATGTTTTGCCTCTTTGCTTCATCCTCACT----TCTAATATTTCTTTTGTG 1002
          ||| |||  |||||||||||||||  ||||||||||||||      ||||||||||||
Db 113933 CCTCACAAGGCAATGATATTGTGCCTTTTTCTTTTCCTACCATCTAATAAGGCTTATGTG 113992

Qy   1003 GAGCTTAGGGTAGGGAATGAAATGAAAAAGGAAGCATACTTGCATTGCATATGTTACTAAGTCAAA 1062
          |||||| |||||||||  ||| |||| |||||| ||||||||||||||||  |||||
Db 113993 GAGTTCAGGGTATGGAAAAGAGTAGAGCATTACTTGTATCACATATTGCAAAGTCAAA 114052

Qy   1063 AACCAAATCTGAGGAGAAGCAAGTCATACAAATCGATC-AAGATGTGCAAGTGTGTGGAT 1121
          |||| |||| ||| ||||||  ||||||||||| ||| |||||||||| |||||
Db 114053 TAATTGATCCACGGAAATACAAGACATACTTTCTGGTCAAAAATGTGCATGTGTGTATGTGT 114112

Qy   1122 ATGTGGATT 1130
          ||| |||
Db 114113 ATGTGGAAT 114121
```

RESULT 11
AP004663
LOCUS   AP004663   144294 bp   DNA   linear   PLN 03-DEC-2003
DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 8,
            PAC clone:P0410E02.

```
ACCESSION    AP004663
VERSION      AP004663.3  GI:38637101
KEYWORDS     .
SOURCE       Oryza sativa (japonica cultivar-group)
  ORGANISM   Oryza sativa (japonica cultivar-group)
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
             Ehrhartoideae; Oryzeae; Oryza.
REFERENCE    1
  AUTHORS    Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE      Oryza sativa nipponbare(GA3) genomic DNA, chromosome 8, PAC
             clone:P0410E02
  JOURNAL    Published Only in Database (2002)
REFERENCE    2  (bases 1 to 144294)
  AUTHORS    Sasaki,T., Matsumoto,T. and Yamamoto,K.
  TITLE      Direct Submission
  JOURNAL    Submitted (23-JAN-2002) Takuji Sasaki, National Institute of
             Agrobiological Sciences, Rice Genome Research Program; Kannondai
             2-1-2, Tsukuba, Ibaraki 305-8602, Japan
             (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
             Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT      On Dec 2, 2003 this sequence version replaced gi:31096587.
             Genes were predicted from the integrated results of the following:
             GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH
             (http://www.softberry.com/), GeneMark.hmm
             (http://opal.biology.gatech.edu/GeneMark/), GlimmerM
             (http://www.tigr.org/tdb/glimmerm/glmr_form.html), RiceHMM
             (http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor
             (http://bioinformatics.iastate.edu/cgi-bin/sp/sp.cgi), sim4
             (http://globin.cse.psu.edu/html/docs/sim4.html), gap2
             (http://www.tigr.org/software/glimmerm/), BLASTN and BLASTX. The
             genomic sequence was searched against NCBI NonRedundant Protein
             database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA
             sequence database at RGP or DDBJ. Protein homologies of the coding
             regions were searched against NCBI NonRedundant Protein database
             with BLASTP. ESTs represent the identified cDNA sequences using
             BLASTN with the corresponding DDBJ accession no. and RGP clone ID.
             Full-length cDNAs represent the identified cDNA sequences using
             BLASTN with identity or significant homology to a protein is
             A gene with identity or significant homology to a protein is
             classified based on the protein name to indicate the homology level
             such as same name, 'putative-' and '-like protein'. A gene without
             significant homology to any protein but with full-length cDNA or
             EST homology (covering almost the entire length of partial
             sequence) is classified as an 'unknown' protein. A gene predicted
             by two or more gene prediction programs is classified as a
             'hypothetical' protein according to IRGSP standard. A gene
             predicted by a single gene prediction program is also classified as
             a probable 'hypothetical' protein and is included as a
             miscellaneous feature of the sequence.
             The orientation of the sequence is from T7 to SP6 of the PAC clone.
             This sequence of P0410E02 clone has an overlap with OJ1484_G09
             (DDBJ: AP003913) clone at 5' end and with OSJNBa0091C18 (DDBJ:
             AP005389) at 3' end.  Detailed information on overlap and assembly
             quality together with annotation of this entry is available at
             http://rgp.dna.affrc.go.jp/GenomeSeq.html.
FEATURES             Location/Qualifiers
     source          1..144294
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="8"
                     /clone="P0410E02"
     gene            complement(5113..5364)
                     /gene="P0410E02"
     misc_feature    complement(5113..5364)
                     /gene="P0410E02.1"
                     /note="hypothetical ORF
                     predicted by GENSCAN"
     gene            10557..13529
                     /gene="P0410E02.2"
```

```
     mRNA            join(10557..10801,11199..11282,12931..13013,13127..13529)
                     /gene="P0410E02.2"
                     /note="supported by full-length cDNA(s): AK061882"
     CDS             join(10660..10801,11199..11282,12931..13013,13127..13246)
                     /gene="P0410E02.2"
                     /note="contains EST(s):
                     D22626(C0623),C27376(C51740),AU100814(C51740)
                     contains full-length cDNA(s): AK120339,AK061882,AK120022"
                     /codon_start=1
                     /product="putative ethylene-responsive transcriptional
                     coactivator"
                     /protein_id="BAD03357.1"
                     /db_xref="GI:38637102"
                     /translation="MAGIGPIRQDWEPVVVRKKAPTAAAKKDEKAVNAARRSGAEIET
                     MKKYNAGTNKAASSGTSLNTKRLDDDTESLAHERVSSDLKKNLMQARLDKKMTQAQLA
                     QMINEKPQVIQEYESGKAIPNQQIIGKLERALGTKLRGKK"
     gene            15250..15480
                     /gene="P0410E02.3"
     mRNA            <15250..>15480
                     /gene="P0410E02.3"
                     /note="start and end point are not identified"
     CDS             15250..15480
                     /gene="P0410E02.3"
                     /note="predicted by GlimmerM etc."
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="BAD03358.1"
                     /db_xref="GI:38637103"
                     /translation="MTLFTSSTSDFRILLCGLLQGGDGETTLAVEVGNQRRREKLRT
                     LGGVCGATDHFACFTMAHGANAGASCLWLCLL"
     gene            16776..19627
                     /gene="P0410E02.4"
     mRNA            join(16776..17137,17808..17871,18919..19627
                     /gene="P0410E02.4"
                     /note="supported by full-length cDNA(s): AK103736"
     CDS             18941..19375
                     /gene="P0410E02.4"
                     /note="contains EST(s): AU055902(S20138),AU055903(S20138)
                     contains full-length cDNA(s): AK103736"
                     /codon_start=1
                     /product="unknown protein"
                     /protein_id="BAD03359.1"
                     /db_xref="GI:38637104"
                     /translation="MEEDSVISNGGREEVVVANGSGGGVGGTARSSGGGGKVVQVL
                     QRNFGEVQGILEQNRVIQEISQNHEARDADGLTRNVALIRELNTNIARVVDLYANLS
                     GSFSRSVTAAASANNTNATNSTSPSAAAAKASKRSRATDTE"
     gene            20343..20591
                     /gene="P0410E02.5"
     misc_feature    20343..20591
                     /gene="P0410E02.5"
                     /note="hypothetical ORF
                     predicted by GlimmerM
                     this category is not included in IRGSP standard"
     gene            complement(22869..27717)
                     /gene="P0410E02.6"
     mRNA            complement(join(22869..23358,25784..26109,27460..>27717))
                     /gene="P0410E02.6"
                     /note="start and end point are not identified"
     CDS             complement(join(22869..23358,25784..26109,27460..27717))
                     /gene="P0410E02.6"
                     /note="predicted by GENSCAN etc."
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="BAD03360.1"
                     /db_xref="GI:38637105"
                     /translation="MRAGTAGCGSCFGCGGGARAAGGGRGARKAMTGGPHPSARAAGG
                     PACQRRAARGEEPMGRRRGERGREKWAEPAQERKGGKKTSFRAWHDGPDTSPPRFPPL
                     PPRRHATSPSPASLLKSSAAPPDSASAPPLCLLQSPTPVGQRRYASGRRQAPIGRL
                     PRRLGPGAQSGGRPHPLEREREEALKRRRHSAGGRSEQSIRLDLGMMEEDSFINNDVG
                     EEVVVANGGDGEGVGGMARSSDDGDGGGGNVVVDDPPLSLGEVRGILEQNHTLIQEISQ
                     NHKARDADRLTRNVALIRDLNTNIARVVDLYANLTGSSHSHTTSDVASATNTNATNSS
                     SPSAAAATAEVSKQPHTTDTDAK"
     gene            complement(27899..28693)
```

```
mRNA          /gene="P0410E02.7"
              complement(join(<27899..28364,28416..28527,28690..>28693))
              /gene="P0410E02.7"
              /note="start and end point are not identified"
CDS           complement(join(27899..28364,28416..28527,28690..28693))
              /gene="P0410E02.7"
              /codon_start=1
              /product="pri-like protein"
              /protein_id="BAD03361.1"
              /db_xref="GI:38637106"
              /translation="MGRGRADRVHGSGAHASASRSTVDRAHAADGGRRRGSHAREGAP
              EGSLAGTADGRGTRPRPDGHRRRPARRSGTRERGGKGEANGAVFGSPRSTATTGTAA
              EAEEGSGAVRDDEDDGAPTVGERNGGADGVDDDAAKPKEATPRGEEVRSDGGGEPELV
              GDGGERGKRRERDSDGSESERRTAETEERGTGNN"
gene          join(29395..29559,30563..31192)
              /gene="P0410E02.8"
misc_feature  join(29395..29559,30563..31192)
              /gene="P0410E02.8"
              /note="polyprotein-like"
.gene         31492..33217
              /gene="P0410E02.10"
repeat_unit   31492..33217
              /gene="P0410E02.10"
              /note="5' terminal repeat"
              /rpt_type=terminal
gene          complement(join(33334..33541,35291..35385))
              /gene="P0410E02.9"
misc_feature  complement(join(33334..33541,35291..35385))
              /gene="P0410E02.9"
              /note="hypothetical ORF
              predicted by GENSCAN
              this category is not included in IRGSP standard"
gene          join(35567..39070,49891..50916)
              /gene="P0410E02.10"
misc_feature  join(35567..39070,49891..50916)
              /gene="P0410E02.10"
              /note="probably inactive due to insertion of
              retrotransposon in CDS
              pseudogene, polyprotein"
              /pseudo
gene          39661..42286
              /gene="P0410E02.11"
repeat_unit   39661..42286
              /gene="P0410E02.11"
              /note="5' terminal repeat"
              /rpt_type=terminal
gene          42868..44271
              /gene="P0410E02.11"
misc_feature  42868..44271
              /gene="P0410E02.11"
              /note="probably inactive due to including frameshift(s) in
```

```
Qy     533 ATAAATCACCAAAACTCATGGAGCTTGCTAGTTATAAAACCTCTAATTCTAAGTTGGT-GT 591
           ||| |||||||| || ||  || | ||| ||||| ||||| ||  |  |  |||||||  |
Db  117165 ATATTTCACCAACACTGGTTGGAAATGATTAGTAGTAATAAACCCTACCACTAAGTTGGATACC 117224

Qy     592 TTATTTGAGTGGATTTTGTGTGAAAGTTGGTGGTTAGAAATAGGAGTTAAGGACCGCCAA 651
           | || | | ||| |||  ||  ||| ||| |  ||| |||| | |   ||||   | | |
Db  117225 ATATTTTATTTCATTTATTGCAAGTGTCAAGTATTGACGGTCAAATATTATGATTTAAGTACCGTCA 117284

Qy     652 CAAGATCCCCACACTTAGCCCCTTTGCTCATCCTCGAGTAAAGTTCAAGGACTAAGGTGG 711
           ||| || || ||||||| ||  |||||| | |  |||||| ||  | | ||| ||  |||
Db  117285 ACACACACCCCACACTTAACCTTTGCTCCGTCCCGAGTGAAGGCTATGGACTCTCACAAG 117344

Qy     712 AACATCTCCTCAAATGGTACGATGCCTGCATATAAGTCATATTATTCCAAGCCTCCACTATACAT 771
           | | |     |  |  | | ||| ||   || ||  || || ||||||| ||| | |  ||  ||
Db  117345 AGTTATGACTTCCTTACTAATGATCATCGCAAACGAATTATTCCAAAGCCATCATCGTGTACTT 117404

Qy     772 GTGAACTTTGAAGTGTCTACCACGCCATCTTGGGTGGTTGAGAAATGGAACAGATCAGAA 831
```

```
Db  117405 GTGGTCTTTGTAGTGTCTACCATGCGATCGTGAGAAATTGAGAAATTAGAACACATTCTTGAC 117464
           ||| |||||| ||||||||||| |  ||| ||||||||||||||| |||||||||| |||||
Qy     832 TCCAGTCATCTTTACCTCCTTGCTTAGATAACTTGGGTTTTGTAAGGTTTTCAAATTT 891
           | || ||| ||  ||||  |||||  ||||| ||| ||||||| |||||||  |||||
Db  117465 TCAAGTCACCCTAATTTTCATGAAGA-ACAACTTAGGGGTTTTCAGTAACACAAAATT 117523

Qy     892 AAAACATAGTCTTGTCTCCTCAAATGATTCTCTCATATAGCTCAATGTGTATGTTTCTCA 951
           | | ||| || ||| |||||||| ||||||||||  ||| || |  |||||| ||| |||
Db  117524 ATAACTTGCTTTGCTCTCCTCAATTGATTCTCTCAATCAG-TCAAGGTGTATGAATTCCTCA 117582

Qy     952 CCAAGGCAAATGTTTGCCCCTTTCATCCTACTTCTCAATATTTCTTTTGTGGAGCTTAGG 1011
           ||||||| ||| | |||| |||    ||| || |||||||||| || ||||||| ||||
Db  117583 CAAAGGCCATATGATTGTGCCTT----TCCTTTTCCTACCAAGGCTTATGTGGAGCTTCAGG 117638

Qy    1012 GTAGGGAATGAAAAGGAAGCATACTTGCATATGTTACTAAGTCAAAAAACCAAATC 1071
           |||  |||| |||| | |||||| ||||||  |  || |||||||  ||   |||
Db  117639 GTATGGAAAAGAGTAGAGCATTACTTGCATCACACATATACTGCAAAGTCAAATAAAGGATC 117698

Qy    1072 TGAGGAGAAGCAAGTCATACAATCTGATCAAG-ATGTGCAAGTGTGTGG 1119
           | ||||||||| || |||||||| |   ||||  |||||| |||| |||
Db  117699 CATGAAAATACAAGACATACTTTCTCGGTCAAGAATGTGCATGTGTATGG 117747
```

```
RESULT 12
AC136521
LOCUS       AC136521               149435 bp    DNA     linear   PLN 08-MAY-2004
DEFINITION  Oryza sativa (japonica cultivar-group) chromosome 5 BAC clone
            OSJNBa0034M22, complete sequence.
ACCESSION   AC136521
VERSION     AC136521.2  GI:47084432
KEYWORDS    HTG.
SOURCE      Oryza sativa (japonica cultivar-group)
ORGANISM    Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 149435)
  AUTHORS   Chow,T.-Y., Hsing,Y.-I.C., Chen,C.-S., Chen,H.-H., Liu,S.-M.,
            Chao,Y.-T., Chang,S.-J., Chen,H.-C., Chen,S.-K., Chen,T.-R.,
            Chen,Y.-L., Cheng,C.-H., Chung,C.-I., Han,S.-Y., Hsiao,S.-H.,
            Hsiung,J.-N., Hsu,C.-H., Huang,J.-J., Kau,P.-I., Lee,M.-C.,
            Leu,H.-L., Li,Y.-F., Lin,S.-J., Lin,Y.-C., Wu,S.-W., Yu,C.-Y.,
            Yu,S.-W., Wu,H.-P., Shaw,J.-F., *McCombie,W.R., *Zutavern,T.,
            *Spiegel,L., *de la Bastide,M., *Muller,S., *Nascimento,L.,
            *Balija,V., *Bell,M., *Miller,B., *Katzenberger,F., *Andrade,M.V.,
            *Dike,S., *o'Shaughnessy,A. and *Palmer,L.
  TITLE     Oryza sativa (japonica cultivar-group) chromosome 5 BAC clone
            OSJNBa0034M22, complete sequence
  JOURNAL   Unpublished
  REMARK    *Lita Annenberg Hazen Genome Center, Cold Spring Harbor Laboratory,
            1 Bungtown Road, Cold Spring Harbor, NY, 11724, USA
REFERENCE   2  (bases 1 to 149435)
  AUTHORS   Chow,T.-Y. and Hsing,Y.-I.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-2002) Institute of Botany, Academia Sinica, 128,
            Section 2, Academia Road, Nankang, Taipei 11529, Taiwan
REFERENCE   3  (bases 1 to 149435)
  AUTHORS   Chow,T.-Y. and Hsing,Y.-I.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAY-2004) Academia Sinica, Institute of Botany, 128,
            Section 2, Academia Road, Nankang, Taipei 11529, Taiwan
COMMENT     On May 8, 2004 this sequence version replaced gi:24635266.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e. phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. This BAC clone was
            sequenced to phase II by the Academia Sinica and was completed to
            phase III by the Cold Spring Harbor Laboratory Genome Center.
            Clone OSJNBa0034M22 overlaps clone P0012A07 (AC134345) from base 1
```

```
Query Match       6.5%;  Score 175.8;  DB 8;  Length 144294;
Best Local Similarity   61.8%;  Pred. No. 4e-32;
Matches 364;  Conservative   0;  Mismatches 217;  Indels   8;  Gaps   5;
```

```
Db              41285 ATGGTGAAT 41293
                      ||| |||
```

```
RESULT 13
AP005389
LOCUS       AP005389           154416 bp    DNA    linear   PLN 03-DEC-2003
DEFINITION  Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 8,
            BAC clone:OSJNBa0091C18.
ACCESSION   AP005389
VERSION     AP005389.3  GI:38637257
KEYWORDS    .
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T. and Katayose,Y.
  TITLE     Oryza sativa nipponbare(GA) genomic DNA, chromosome 8, BAC
            clone:OSJNBa0091C18
  JOURNAL   Published Only in Database (2002)
REFERENCE   2  (bases 1 to 154416)
  AUTHORS   Sasaki,T., Matsumoto,T. and Katayose,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2002) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     On Dec 2, 2003 this sequence version replaced gi:28569997.
            Genes were predicted from the integrated results of the following:
            GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH
            (http://www.softberry.com/), GeneMark.hmm
            (http://opal.biology.gatech.edu/GeneMark/), GlimmerM
            (http://www.tigr.org/tdb/glimmerm/glmr_form.html), RiceHMM
            (http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor
            (http://bioinformatics.iastate.edu/cgi-bin/sp/sp.cgi), sim4
            (http://globin.cse.psu.edu/html/docs/sim4.html), gap2
            (http://www.tigr.org/software/glimmerm/), BLASTN and BLASTX. The
            genomic sequence was searched against NCBI NonRedundant Protein
            database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA
            sequence database at RGP or DDBJ. Protein homologies of the coding
            regions were searched against NCBI NonRedundant Protein database
            with BLASTP. ESTs represent the identified cDNA sequences using
            BLASTN with the corresponding DDBJ accession no. and RGP clone ID.
            Full-length cDNAs represent the identified cDNA sequences using
            BLASTN with the corresponding DDBJ accession no.
            A gene with identity or significant homology to a protein is
            classified based on the protein name to indicate the homology level
            such as same name, 'putative' and '-like protein'. A gene without
            significant homology to any protein but with full-length cDNA or
            EST homology (covering almost the entire length of partial
            sequence) is classified as an 'unknown' protein. A gene predicted
            by two or more gene prediction programs is classified as a
            'hypothetical' protein according to IRGSP standard. A gene
            predicted by a single gene prediction program is also classified as
            a probable 'hypothetical' protein and is included as a
            miscellaneous feature of the sequence.
            The orientation of the sequence is from M13rev to -21M13 of the BAC
            clone. This sequence of OSJNBa0091C18 clone has an overlap with
            P0410E02 (DDBJ: AP004663) clone at 5' end and and with OJ1770_H03
            (DDBJ: AP005298) at 3' end. Detailed information on overlap and
            assembly quality together with annotation of this entry is
            available at http://rgp.dna.affrc.go.jp/GenomeSeq.html.
FEATURES             Location/Qualifiers
     source          1..154416
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="8"
                     /clone="OSJNBa0091C18"
```

```
                to base 64031. The overlap is from base 72767 to base 136858 on
                P0012A07. OSJNBa0034M22 also overlaps clone P0508G08 (AC130726)
                from base 136914 to base 149435. The overlap is from base 1 to
                base 12522 on P0508G08.
FEATURES             Location/Qualifiers
     source          1..149435
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="5"
                     /clone_lib="HindIII"
     misc_feature    5538..5565
                     /note="The sequence is a dinucleotide (TA) repeat in which
                     the exact number of TA pairs is unknown. The overlapping
                     clone P0012A07 (AC134345) shows 31 additional TA pairs
                     compared to that which is represented by the assembly."
     misc_feature    106220..106269
                     /note="The sequence is a dinucleotide (TA) repeat in which
                     the exact number of TA pairs is unknown. Sequences in the
                     region show up to 13 fewer TA pairs than that which is
                     represented by the assembly."
     misc_feature    128275..128340
                     /note="The assembly is covered by high quality sequences
                     derived from a transposed plasmid subclone."
ORIGIN

Query Match      6.5%;  Score 175.8;  DB 8;  Length 149435;
Best Local Similarity  62.1%;  Pred. No. 4e-32;
Matches  378;  Conservative  0; Mismatches 222;  Indels  9;  Gaps  6;

Qy     529 AATGATAAATCACCAAAAACTCATGGAGCTTGCCTAGTTATAAACTCTAATTCTAAGTTTGG 588
           ||||| ||||||||| ||||||||||||| ||||| || |||  || ||||||||||||||
Db   40687 ACTCATATTTCACCAACACTAGTGGAAATGATTAGTAATAAATCTTACCACTAAGTTGGA 40746

Qy     589 T-GTTTATTGGAGTGGATTTGTGTGAAAGTTGGTGTTAGAAATAGGAGTTAAGGACCG 647
           | ||||| |||  ||| |||| || ||||||  || |||||||| ||||| ||||||
Db   40747 TACCATATTTATTTCATTATATGCAGGTATTGCAGGTATTGGTCAAATATTATGATTTAAGGACC 40806

Qy     648 CCAACAAGATCCCCACACTTAGCCCTTTGCTCATCCTCGAGTAAAGTTCAAGGACTAAG 707
           |||||| ||| |||  |||  |||| ||||||||||||||| ||| |||||||| |||
Db   40807 GTCACAACACCCCCACACTTAACCTTTGCCGTCCGAGTGAAGGCTATGGACTATACA 40866

Qy     708 GTGGAACATCTCCTCAAATGGTACGATGCTGCTACGATATAGTTATTCCAAGCCTCACCTAT 767
           ||||| ||||| ||||| |||||| |||||  ||||||| ||||| ||||||||| |||||
Db   40867 ATAGAGTTTTGTCTTCCCTACGAATAATCATGCAAACAAATTATTTAAAGTCATACCTGT 40926

Qy     768 ACATGT-GAACTTTGAAGTTGTCTACCACGCCATCTGGGTGGTTGAGAAATGGAACAGAT 826
           ||||   |||  || ||||| ||||||| || |||||||||||||||||||||||||||
Db   40927 ACTTGTGGGATCTTTGTAGTGTCTACCATGCGATCCTCGGGCAATTGAGAAATGGAATAATC 40986

Qy     827 CAGAATCCAGTCATCTTACCTCTCTTGCTTAGATAACCTGGGTTTTTGTAAGGTTTTCA 886
           |||||||| ||||| ||||  ||||||||||||||||| || || |||||||| ||||
Db   40987 TTGACTCAAGTCACCCTAACTTTTCATGAAGA-ACAACTTGGAGTTTTCACGTAACACA 41045

Qy     887 AATTTAAAACATAGTCTTGCTCCTCCCAAATGATTCTCATAGCTCATCATGTGTATGGTT 946
           || ||||| || ||||||||||| ||||||| ||||||||| |||| || ||||| |||
Db   41046 AAATTATAACTTTGCTTTGCTCCCTCCTATTGATTCTCTCG-ATCACTCAAAGTGTATGATT 41104

Qy     947 TCTCACCAAGGCAATGTTTTGCCTCTTTTCATCCTTACT---TCTAATATTTCTTTTGTG 1002
           ||||| |||||||| ||||||||| ||||||||| |||     ||||||||| ||||||
Db   41105 CCTCACAAAGGCATAGATTGTGCCTTTTTCTTTTCCTACCACCATCCTAATAAGGCTTATGTG 41164

Qy    1003 GAGCTTAGGTGGTAGGGAATGAAAAGGAAGCATACTTGCATTGCATATGTTACTAAGTCAA 1062
           |||||||||||||||| ||||||  ||||||| || ||||||||||||| ||||||||||
Db   41165 GAGTTTCAGGTATGGAAAGAGTAGAGCATTACTTGTATCACATATATTGCAAAGTCAAA 41224

Qy    1063 AACCAAATCTGAGGAGGAGAGCAAGTCATACAATCTGATC-AAGATGTGCAAGTGTGTGGAT 1121
           |||| ||| ||  |||| || | |||||  |||| ||| | ||| |||||||| |||||||
Db   41225 TAATTGATCCACCGAAATACAAGACATACTTTCTGGTCAAAAAATGTGCATGTGTATGTGT 41284

Qy    1122 ATGTGGATT 1130
```

gene   complement(3797..4395)
                /gene="OSJNBa0091C18.1"
misc_feature   complement(3797..4395)
                /gene="OSJNBa0091C18.1"
                /note="probably inactive due to including frameshift(s) in CDS
                pseudogene, Fourf gag/pol protein"
                /pseudo

gene   complement(join(4911..5050,5955..6039))
                /gene="OSJNBa0091C18.2"
misc_feature   complement(join(4911..5050,5955..6039))
                /gene="OSJNBa0091C18.2"
                /note="hypothetical ORF
                predicted by RiceHMM
                this category is not included in IRGSP standard"

gene   complement(7562..8824)
                /gene="OSJNBa0091C18.3"
misc_feature   complement(7562..8824)
                /gene="OSJNBa0091C18.3"
                /note="probably inactive due to including frameshift(s) in CDS
                pseudogene, retroelement"
                /pseudo

gene   complement(10318..10601)
                /gene="OSJNBa0091C18.4"
mRNA   complement(join(<10318..10416,10512..>10601))
                /gene="OSJNBa0091C18.4"
                /note="start and end point are not identified"
CDS   complement(join(10318..10416,10512..10601))
                /gene="OSJNBa0091C18.4"
                /note="predicted by GeneMark.hmm etc."
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="BAD03522.1"
                /db_xref="GI:38637258"
                /translation="MDEVGGRWTFSRANEGYTSCGSVVEMSWQRDGLGTMSRLPSVYR
                GGMWHVRKMCLVGKDVHL"

gene   join(11138..11225,11462..11565,11796..12372,12781..12814,
                13043..13055)
                /gene="OSJNBa0091C18.5"
misc_feature   join(11138..11225,11462..11565,11796..12372,12781..12814,
                13043..13055)
                /gene="OSJNBa0091C18.5"
                /note="hypothetical ORF
                predicted by GENSCAN
                this category is not included in IRGSP standard"

gene   complement(13363..14064)
                /gene="OSJNBa0091C18.6"
mRNA   complement(join(13363..13422,13515..13658,13771..>14064))
                /gene="OSJNBa0091C18.6"
                /note="start and end point are not identified"
CDS   complement(join(13363..13422,13515..13658,13771..14064))
                /gene="OSJNBa0091C18.6"
                /note="similar to Oryza sativa chromosome 4,
                OSJNB0040D15.1"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="BAD03523.1"
                /db_xref="GI:38637259"
                /translation="MGDDELAGVGEKELASGGQMRAGKRRDRDKLMARGLGETEWHGT
                KKIRTPVAYSBSSANWVPNAESSAYSVAYYENFLDEGTDRAAFYDHWQHCRWKRVIRD
                STVLMRNGLHASIHGAETVDLKFTSRKIVQLKNVQPVPTINRNLAMSVEACSTFPFLI
                SAISL"

gene   14636..14968
                /gene="OSJNBa0091C18.7"
misc_feature   14636..14968
                /gene="OSJNBa0091C18.7"
                /note="hypothetical ORF
                predicted by GlimmerM
                this category is not included in IRGSP standard"

gene   complement(18921..22021)

misc_feature   complement(18921..22021)
                /gene="OSJNBa0091C18.8"
                /note="probably inactive due to including frameshift(s) in
                CDS
                pseudogene, copia-type pol polyprotein"
                /pseudo

gene   complement(24160..24510)
                /gene="OSJNBa0091C18.9"
misc_feature   complement(24160..24510)
                /gene="OSJNBa0091C18.9"
                /note="retroelement-like"

gene   complement(27085..28032)
                /gene="OSJNBa0091C18.10"
mRNA   complement(join(<27085..27177,27258..27351,27986..>28032))
                /gene="OSJNBa0091C18.10"
                /note="start and end point are not identified"
CDS   complement(join(27085..27177,27258..27351,27986..28032))
                /gene="OSJNBa0091C18.10"
                /note="predicted by GeneMark.hmm etc."
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="BAD03524.1"
                /db_xref="GI:38637260"
                /translation="MRLSFQAKAQIVIRSSGSMEFVPPADPSMFFPISVGFSTSNTFS
                DVKVPGIRPLKEGSNPPKYSQRVRLVADNYQVV"

gene   complement(28257..28463)
                /gene="OSJNBa0091C18.11"
mRNA   complement(<28257..>28463)
                /gene="OSJNBa0091C18.11"
                /note="start and end point are not identified"
CDS   complement(28257..28463)
                /gene="OSJNBa0091C18.11"
                /note="similar to Oryza sativa chromosome 7,
                OJ1354_H07.17"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="BAD03525.1"
                /db_xref="GI:38637261"
                /translation="MPLGPSCRLAPRPSPNCHVAGGASGEGTQAEPLSMRAAPQLPLA
                AFNVVGADMRYRPTDARQSGMTGL"

gene   28734..29144
                /gene="OSJNBa0091C18.12"
mRNA   <28734..>29144
                /gene="OSJNBa0091C18.12"
                /note="start and end point are not identified"
CDS   28734..29144
                /gene="OSJNBa0091C18.12"
                /note="predicted by GeneMark.hmm etc."
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="BAD03526.1"
                /db_xref="GI:38637262"
                /translation="MLWDGIPSPNGIAPPPQRDSCSLTVVVPPLPRPSPVASTVAAT
                AAAAPVRRRRRYPPLPPRPCPVDTAATVAVPCRRQARPPPPPCPSPATAAIAVPV
                PCRRRRRPHLHGTAAGDRLRVSRCSFPNGMPSYI"

gene   complement(30530..30877)
                /gene="OSJNBa0091C18.13"
mRNA   complement(<30530..>30877)

Query Match   6.5%; Score 175.8; DB 8; Length 154416;
Best Local Similarity 61.8%; Pred. No. 4e-32;
Matches 364; Conservative 0; Mismatches 217; Indels 8; Gaps 5;

Qy   533 ATAAATCACCAAAACTCATGGAGCTTGCTACTTGTTATTAAAATCTAATTCTAAGTTGGT-GT 591
            ||| |||||||| |||| ||| |||| ||| ||||||| ||||||| ||||||| ||||||
Db   5503 ATATTTCACCAACACTGGTGGAAATGATTAGTAATAAACCCTACCACTAAGTTGGATACC 5562

Qy   592 TTATTTGAGTGGATTTTGTGTGAAAGTTGGTGGTTAGAAATAGGAGTTAAGGACCGCCAA 651
            |||||||| |||||| |||||||| |||| |||||| |||| ||| |||| |||| ||
Db   5563 ATATTTTATTTCATTTATATGCAAGTATATGACGGTCAAATATTATGATTTATGATTTAAGTAAGTACCGTCA 5622

```
Qy     652 CAAGATCCCCACACTTAGCCTTTGCTCATCCTCGAGTAAGTTCAAGGACTAAGGTGG 711
               |||||||||| |||||| |||  |||||||||  ||| |||  || ||
Db    5623 ACAACACCCCCACACTTAACCTTTGCTCGTCCGTCCGAGTGAAGGCTATGGACTTCACAACAG 5682

Qy     712 AACACTCTCCTCAAATGGTACGATGCCTGCATATAAGTATTCCAAGCCTCACCTATACAT 771
               || ||| |||||||||||||||| |||| |||  |||||| ||||||||||  |||
Db    5683 AGTTATGACTTCCTTACTAATGATCATGCAAACGAATTATCCAAGACCATATCTGTACTT 5742

Qy     772 GTGAACTTGAAGTGTCTACCACGCCATCTTGGGTGGTTGAGAAATGGAACAGATCAGAA 831
               |||||||||||||||| |||| |||||| ||| ||||||||||||||||||||||||||
Db    5743 GTGGTCTTTGTAGTGTCTACCACCATGCGGATCGTGAGAAATTAGAACAATCTTGAC 5802

Qy     832 TCCAGTCATCTTACCTCTCTGCTTAGATAACTTGGGTTTTTCAAATTT 891
               ||||||||| ||| ||||| |||  |||| ||||||||||| ||||||
Db    5803 TCAAGGTCACCCTAATTTTCACATGAAGA-ACAACTTAGGGTTTTCAAGTAACACAAAATT 5861

Qy     892 AAAACATAGTCTTGCTCCTCAAATGATTCTTCCATATAGCTCAATGTGTATGGTTTCTCA 951
               ||||||||||||||| || |||||||||||||| ||||||||||||||||||||| ||||
Db    5862 ATAACTTTGCTTGCTCCTCAATTGATTCTCTCCAATCAG-TCAAGGTGTATGATTCCTCA 5920

Qy     952 CCAAGGCCAATGTTTTGCCCTCTTCATCCTCTTCTAATATTCTTTTGTGGGAGCTTAGG 1011
               || |||||| ||||| |||||| ||| |||       |||||||||||||||||||||||
Db    5921 CAAGGCCATATGATTGTGCCTT----TCCTTTTCCTACCAAGGCTTATGTGGAGCTCAGG 5976

Qy    1012 GTAGGGAATGCAAAAGGAAGCATATCCATTGCATATGTTACTAAGTCAAAAACCAAATC 1071
               ||||||||||||||||||||| |||| |||||||||||||||||||||||||| |||||
Db    5977 GTATGGAAAAGAGTAGAGCATTACTTGCATCACATATACTGCAAAGTCAAATAAAGGATC 6036

Qy    1072 TGAGGAGAAGCAAGTCATCACAAATCTGATCAAG-ATGTGCAAGTGTGTGG 1119
               |||||||| ||||| |||||| ||| |||||||| |||||||||||||||||
Db    6037 CATGAAAATACAAGACATACTTTCTGGTCAAGAGAATGTGCATGTGTATGG 6085
```

```
RESULT 14
LOCUS       OSJN00040              127717 bp    DNA       linear   PLN 10-FEB-2004
DEFINITION  Oryza sativa genomic DNA, chromosome 4, BAC clone: OSJNBb0050N09,
            complete sequence.
ACCESSION   AL606603
VERSION     AL606603.3  GI:32488200
KEYWORDS    HTG.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Feng,Q., Zhang,Y., Hao,P., Wang,S., Fu,G., Huang,Y., Li,Y., Zhu,J.,
            Liu,Y., Hu,X., Jia,P., Zhang,Y., Zhao,Q., Ying,K., Yu,S., Tang,Y.,
            Weng,Q., Zhang,L., Lu,Y., Mu,J., Lu,Y., Zhang,L.S., Yu,Z., Fan,D.,
            Liu,X., Lu,T., Li,C., Wu,Y., Sun,T., Lei,H., Li,T., Hu,H., Guan,J.,
            Wu,M., Zhang,R., Zhou,B., Chen,Z., Chen,L., Jin,Z., Wang,R.,
            Yin,H., Cai,Z., Ren,S., Lv,G., Gu,W., Zhu,G., Tu,Y., Jia,J.,
            Zhang,Y., Chen,J., Kang,H., Chen,X., Shao,C., Sun,Y., Hu,Q.,
            Zhang,X., Zhang,W., Wang,L., Ding,C., Sheng,H., Gu,J., Chen,S.,
            Ni,L., Zhu,F., Chen,W., Lan,L., Lai,Y., Cheng,Z., Gu,M., Jiang,J.,
            Li,J., Hong,G., Xue,Y. and Han,B.
  TITLE     Sequence and analysis of rice chromosome 4
  JOURNAL   Nature 420 (6913), 316-320 (2002)
   MEDLINE   22337377
   PUBMED    12447439
REFERENCE   2
  AUTHORS   Fu,G., Wang,S.Y., Ren,S.X., Lv,G., Lin,W., Gu,W.Q., Zhu,G.F.,
            Tu,Y.F., Jia,J., Yin,H.F., Zhang,Y., Cai,Z., Chen,J., Kang,H.,
            Chen,X.Y., Shao,C.Y., Sun,Y., Hu,Q.P., Zhang,X.L., Zhang,H.,
            Wang,L.J., Ding,C.W., Sheng,H.H., Gu,J.L., Chen,S.T., Ni,L.,
            Zhu,F.H., Han,B., Feng,Q., Huang,Y.C., Li,Y., Zhu,J.J., Zhao,Q.,
            Zhang,Y., Liu,Y.L., Mu,J., Yu,Z., Chen,L., Fan,D.L., Weng,Q.J.,
            Li,C., Li,T., Zhang,Y., Hu,S.L., Liu,X.H., Lu,T.T., Zhang,Y.J., Lu,Y.,
            Zhou,B., Chen,Z.H., Hao,P., Zhang,L., Wu,M., Zhang,R.Q., Guan,J.P.
            and Hong,G.F.
  TITLE     Direct Submission
```

```
JOURNAL    Submitted (08-SEP-2001) Han Bin, National Center for Gene Research,
           Chinese Academy of sciences, 500# Cao Bao Road, Shanghai 200233,
           CHINA. E-mail enquiries: bhan@ncgr.ac.cn. Clone requests:
           bhan@ncgr.ac.cn
REMARK     Oryza sativa japonica (nipponbare) genomic DNA, chromosome 4, BAC
           clone: OSJNBb0050N09.
COMMENT    On Jul 9, 2003 this sequence version replaced gi:21912461.
           Web site: http://www.ncgr.ac.cn
           ------------- Summary Statistics
           Assembly program: phrap

           This is a complete sequence.
           Genes were identified by a combination of several methods: Gene
           prediction programs including Fgenesh (http://www.softberry.com/),
           genscan (http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM
           (http://genemark.biology.gatech.edu/GeneMark/), tRNAscan-SE (Sean
           Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/), searches of the
           complete sequence against NCBI none redundant protein database (nr)
           (ftp://ncbi.nlm.nih.gov/blast/db) and the EST database at NCGR.
FEATURES            Location/Qualifiers
     source          1..127717
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /sub_species="japonica"
                     /db_xref="taxon:39947"
                     /chromosome="4"
                     /clone="OSJNBb0050N09"
                     /clone_lib="CUGI-OSJNBb"
     gene            16286..17314
                     /gene="OSJNBb0050N09.1"
     CDS             16286..17314
                     /gene="OSJNBb0050N09.1"
                     /codon_start=1
                     /protein_id="CAE02950.1"
                     /db_xref="GI:32488201"
                     /db_xref="UniProt/TrEMBL:Q7X7R8"
                     /translation="MGVQEEAALAAAVVKGKRSKRQRAHAAVVPIAASATAEEEESME
                     SSLSLSGGAAAEGSSTTSPLLTTTAARGDEAVSGCVTEEEDMALCLMLLASGGHGER
                     APDAEAVAKEAKFSRRPADGAGAGEFVYECKTCSKCFPSFQALGGHRTSHKKPRLV
                     APPAAMEPAADDKIKPTIPETAAAAAEEKPKPKSPPRPPASRPIATDPTVLAIPVIPK
                     QEVIDATNSAAIASVSKQPRVHECSICGAEFASGQALGGHMRHRPLIPASASSAVVS
                     VLDAVDAPRQKEKSLLEIDLNMPAPCDDATAETTTSSAATSPAFAFAVSDRSPLLVP
                     AALVGCHY"
     gene            complement(29123..30334)
                     /gene="OSJNBb0050N09.2"
     CDS             complement(29123..30334)
                     /gene="OSJNBb0050N09.2"
                     /codon_start=1
                     /protein_id="CAE02951.2"
                     /db_xref="GI:38345136"
                     /db_xref="UniProt/TrEMBL:Q7X7A0"
                     /translation="MRAPPPSPAVARTGFSDLQPDLVGEIHGRLSFLDRLAFVAVSSA
                     ASRDALKPEPPWLVLPGETPETATVFSLADRRSATVRASGPAMRGHVIIGSFGGWIVT
                     ADERGRMRLANPVTGDQDLPAITTIPFVNATSPWGHHFIMDMEPFVHIRYRGDHESW
                     PHPYGTPTHTAGDIRLMFYRKVVLSASPRPGDYAAMLLLDSYFGAPAFATAGDGRWRV
                     APSRDGVEDAIHHGGKPLSVYTGIVEAWERDGVSGFTGKVTTPISCPPQRRKYLA
                     AAPDGRIMVLKLVNTNGGVKKGYFEVQVFDEMTQRWEAAEDIGELAILLVGVNSSVCVST
                     AKHPELKGGCVYTDDEIGKAWLRREYGYSNSKPNVGVYSLKDGKEMSIPGLGEHLSW
                     PPPAWFTPSFR"
     gene            complement(39012..41270)
                     /gene="OSJNBb0050N09.3"
     CDS             complement(39012..41270)
                     /gene="OSJNBb0050N09.3"
                     /codon_start=1
                     /protein_id="CAE02952.1"
                     /db_xref="GI:32488203"
                     /db_xref="UniProt/TrEMBL:Q7XQW6"
                     /translation="MAGDGWVMDIIHWTLLLGGDAMGGIWSSWSTEILLGLSFVAQLV
                     LTVTAGFRWRGAGSRMRCVIWFSYVSADYVATTALGNLSISRTAGSRRLAAFWAPFFL
                     LHLGGPDSVTAYELEDNQLSARHVLEILIRVAGAVYIVYKSTGSSWALIPASWLMLFV
                     GVAKYAEKTMALRRANLANVRRTVERERLQRRSRTTKANLSFAGDDDEGGLLMKAH
                     TLFPICKNSMVDDSSVETASNTDDAAIVHAKETLFREENYKNVFRVMEMELSLMYDFLY"
```

TKAAVIHTWHGYAIRAVSPVFTAVSLVLVELSNVAGHHRRSRSDVVITRVLLVATFLLET
LSLLRALASSWTGFWLDRELRPVWGCGWIRHEVLCRSRWTWLRRQVASIGRLAGAKDH
RRWCGKWGQLSVLQLIITGSASEQEDRSWDKECETYSKEKTIVVPQDVKELFFRRLLG
QLIDLRKRMKADTGTETELRTMVANMSKRGQLTLQNYDLWNQLRWSLGDELQLGLIT
WHIATDIYLSQSVKAIVAAVEDDAVLARWLMGIRTLSNYMVHLLAVRPDMLPGLVTRK
LFELTCDDLARVWSKHGTSTSVGAAGSDLESSSSSPRNICRLRDLMRVSPKTIEQQN
KLAGMLIKQWEWDRKQPESGAVELNKYLSRGIELAKKLLHLEVSRKDIDKVLQVILEVW
VEMLFYAGYRCSKESHAKQLSQGGELTTIVWLMAEHVGLFLVNKTSKGAEEDYWNTRK
RRYSRQPASQNV"

gene   complement(46601..47527)
            /gene="OSJNBb0050N09.4"
CDS   complement(46601..47527)
            /gene="OSJNBb0050N09.4"
            /codon_start=1
            /protein_id="CAE02953.2"
            /db_xref="GI:38345137"
            /db_xref="UniProt/TrEMBL:Q7XQW5"
            /translation="MTRRGTFCVLFPRRYSKGKLKNILSIPRSSWCQPKSRNSRYRLP
EMRAPPLPLPAATTGFSDLQPELMGEIHGRLSFLDHLAFAAASSAATRDAFKPEPPW
LVIPGDTPGTATVFSTRVAPSLDSVEDAIHHEGKFLSVTYTGIVEAWERDGVSGEFTS
NVVTTRMADGGDDDHHRKYLAAAPDGRLMIVLKNTKRMKQQNHFKVQVFDEMTQRWEAA
EDIGELPILVGVNSSLCVSTAKHPEINAGCVYYTDDDMWEASMRRDFWCSNRKPNVGA
YSLKDGKAMSIPGLGEHLSWPPPAWFTPSFPR"

gene   complement(50686..54208)
            /gene="OSJNBb0050N09.5"
CDS   complement(join(50686..50986,51066..51236,51319..51542,
            51588..51922,52207..52267,53008..53139,53221..53682,
            53738..54208))
            /gene="OSJNBb0050N09.5"
            /codon_start=1
            /protein_id="CAE02954.2"
            /db_xref="GI:38345138"
            /db_xref="UniProt/TrEMBL:Q7XQW4"
            /translation="MEKWFTFPAEAKDRLKRKAFQKMGKAWKNWKSKLFTEFVNQPGN
HTPFDEYPQITEAVWEEFCSLKNTQEFRESSEAHRLLQQRNEHPHWLGTAGYIGKEAI
WSQEDAAAAVANVPAPFSDISEQRARNWAARGKVNPDGSVTFENKSDAVVYQELLGL
VVEQASQSEVESALKRREDDILAKVLGTKEHPGRTRGIGSDVRWKHRLPQYSSQYRKQ
KVSKEERDARLKAELKAEVIEELEASMNARVEERVNKVLADMNIPRGTTPAVHPTPRA
QHDASPSQHRSSCAFTEVPAPGLAVAPLAAVDHIESVAQCVLLARVHPTFAPEVAEGM
AFKPSVTDKVHGADLLAGYAKALGALEARSETGCARVRWDPLQWPKEDIVVKMTRP
SRPMELTPPKSKLSIEAPRGPALSVPHSPGADMDLADIAQSLAPIKTTRKADSSPPL
VKGQKRERGKGKVGELAPEPKRGKAATSMPVSKAGKVDNVLAVMGIACRELHKQYMEL
SNAKRKIRESSIVGHHDHQPFLSSPAYITIGFDDLPDLFRIRRKLDTGLLKCYSLLCWI
ESRRLGNVGFLDPSMVNEVNLRQSFTEVVDYVNRCLWAHQDKEYIMCAHRNQERHWIL
LVIVPKWSRVTYLNSNKSKDYDFTEITKALNMAWDPVVEKGGRHKEGKDELYHDTKFA
CAQQIRDQCGFHVCHNMSTFLREVKDFDPEVVANGE"

gene   complement(58336..61955)
            /gene="OSJNBb0050N09.6"
CDS   complement(join(58336..59277,59410..59639,59677..60276,
            60641..60716,60765..60878,61155..61422,61664..61802,
            61841..61955))
            /gene="OSJNBb0050N09.6"
            /codon_start=1
            /protein_id="CAE02955.2"
            /db_xref="GI:38345139"
            /db_xref="UniProt/TrEMBL:Q7XQW3"
            /translation="MTVTFDGSLGLPSSVGHPAVARHRYPLTCDGPEGKAVTGIPTPV
TGRGIMSRYRWWSATCDGFSFRIVTGRSIPLTGNSIEARQRMIVDGCTMRIIRLPSIE
RGVTEFVTFADNRDQSRMSMRMLCPCRDCKNEQMIEDSDEVHAHLIMNGFMKKYTCWT
KHGEQEAPDVAADEDKSFTELLELLKDLPBGNNLPQTTYEAKQVLCPLGLEEYADLD
VCPVYGASRYKRATSEGEGRRPQPGNDILLDFFERLMNEGTRTWDAYAQEYF
NLHVMLFCTINDYPALGNLSGGTVKGKWACSECMEETRSKWLKHSHKTVYMGHRRFLP
RYHPYRNMKQNFNGHRDTAGPPAELTGTEVHKLVMGIINEFGKKRKVGKTKEKSTSKE
KTEEHVEQKTKERSMWKKKSIFWRLPVWKDLEVRHCGDACREBECTVKDGLNARRDL
EEMGVRSELHPITTESGRVYLPPACYTLSKEEKIDLLTCLSGIKVPSGYSSRISRLVS
LQDLKLKIIDPPEGLDELHAELVRTLCHLEMYPPTFFDIMEHLPVHLVRQTKCCGPAF
MTQMYPCERYLGLLKGYVRNRSHPEGSDVIVSYTSEWRTMQGSQPEWSATTWQGYDINGYIFHTVKQD
VRNKHMSSFNEWLKNRIARLQNLSSETLQWLSQGPEWSATTWQGYDINGYIFHTVKQD
SKCTVQNSGLRIEAASDSGRRDQYYGRVEQILBLDYLKFKVPLFRCRWVDLRNVKVDN
EGFTTVNLANNAYKDEPFVLAKQVVKCST"

gene   complement(63715..65010)
            /gene="OSJNBb0050N09.7"

CDS   complement(63715..65010)
            /gene="OSJNBb0050N09.7"
            /codon_start=1
            /protein_id="CAE02956.2"
            /db_xref="GI:38345140"
            /db_xref="UniProt/TrEMBL:Q7XQW2"
            /translation="MAGTAFSDLPPELMAKIHDNLAFLDRLAFAAAVSAPSRDALKPE
PPWLVLPGETQETTRVFSLADRSATVRAPDPAWRDHVIIGSSGWIATADERGRMRL
ANPVTGEQGDLPAIATIPFLKADQGGRYFVLIMEPFVQLRYRGELESWPLRPHPYGTF
TLTDGDMRRWFYRKVVLSASPRPGDYAAMLLGNYFGTPAFATAEDGRWRVTPSRDGV
EDAIHHKGKFLSVTYTGTVEAWERDGVHGEFTSKVVTTRMANGGDHRHCSKYLAAAPD
GRLMIVLNAKRYKDLFELQVFDEKTQRWEAAEDVGDLALLVGINSSLCVSTTEHPEL
KAGCVYYTDDKIGKASLRRGARCSSWYHRGDESHNDDSDRNVAMYSLKDGTAESIPEL
GEHLSWPPPACSYLPSHDRCAGLRFLDQRTPESIFRVYI"

gene   65392..66167
            /gene="OSJNBb0050N09.8"

Query Match   6.5%;   Score 173.6;   DB 8;   Length 127117;
Best Local Similarity   60.9%;   Pred. No. 1.4e-31;
Matches   368;   Conservative   0;   Mismatches   229;   Indels   7;   Gaps   5;

```
Qy        533 ATAAATCACCAAAACTCATGGAGCTTGCTAGTATAAACTCTAATTCTAAGTTGGT-GT  591
              ||| ||||||| ||| |||||||||||||||||||||||||||||| |||||| ||
Db     108876 ATATTTCACCAACACTAGTGGAAATGATCAGTAAATACTACCTACCTAAGTTGGATACC  108935

Qy        592 TTATTGGAGTGGATTTGTGTGAAAGTTGGTGGTTAGAAATAGGAGTAAGGACCGCCAA  651
              ||| ||||| || |||| |||||| ||||||||| |||||||||||||||||||||||
Db     108936 ATGTTTTATTTCATTATATGCAGGTATTGACAGGTCAGATTATGAAGGACCGTCA  108995

Qy        652 CAAGATCCCCACCACTTAGCCCTTTGCTCATCCTCCGAGTAAAGTTCAAGGACTAAGGTGG  711
              ||||||||||  ||||||||||| |||||| ||||||||||||||||||||||| |||
Db     108996 ACAACACCCCCACACTTAACCTTTACTCGTCCCGAGTGAAGGCTATGTATTCTACAATAG  109055

Qy        712 AACATCTCCTCAAATGTACGATGCGTCGCATATAAGTATTCCAAGCCTCCACCTATACAT  771
              |||||||||| ||||||||| ||| ||||||| |||| ||||| |||||||||||
Db     109056 AGTTATGACTTCCTTACTAATAATTACGCAAACAAATTATTCAAAGCCATACATGTACTT  109115

Qy        772 GT-GAACTTTGAAGTGTCTACCACGCCATCTTGGGTGGTTGAGAAATGGAACAGATCAGA  830
              || |||||||||| || || |||| || ||||||||||||||| ||||||||||||||
Db     109116 GTGGGGTCTTTGTAGTGTCTACCATGCGATCCTGGGCAATTGAGAAATGGAACAATCTTGA  109175

Qy        831 ATCCAGTCATCTTTACCTCCTTGCTTAGATAACTTGGGTTTTTGTAAGGTTTTCAAATT  890
              ||||| ||||||| ||| || ||||| ||| ||  |||| ||||| |||||||| |||
Db     109176 CTCAAGTCACCCTAACTTTTCATGAAGAACAACTTGGGGGGTTTTCAAGTAACACAAAAT  109235

Qy        891 TAAAACATAGTCTTGCTCCTCAAATGATTCTCATATAGCTCAATGTGATGGTTTCTC  950
              ||||| |||||||||| |||| ||||||||||| ||| |||| |||||||||||||||
Db     109236 TATAACTTTGCTTGCTCCTCCATTGATTCTCTCG-ATCACTCAAAGTGTATGATTCCTC  109294

Qy        951 ACCAAGGCAATGTTTGCCTCTTTTCAT---CCTACTTCTAATATTTCTTTTGTGGAGCT  1007
              ||||||| |||||| |||||||||| ||   |||| |||| ||||| ||||||||||||
Db     109295 ACAAAGGCATATGAATGTGCCCTTTCTTTTTCCGCCATTTAATAAAGCTTATGTGGAGCT  109354

Qy       1008 TAGGGTAGGGAATGAAAAGGAAGAGCATACTTGCCATATGCATAGTCAAAAACCA  1067
              ||||| ||||| |||||||||||||||||||||||||||||||| |||||||| |
Db     109355 CAGGGTATGAAAAGAGAGAGTAGAGCATTACTTGCCATCATATATTGCAAAGTCAAATAATG  109414

Qy       1068 AATCTGAGAGAAGCAAGTCATCAAATCTGATCAAG-ATGTGCAAGTGTGTGGATATGTG  1126
              ||||||||| |||| ||||||| |||| |||| ||| |||||||||||||||||||||||
Db     109415 GATCCATGGAAATACAAGACATACTTCGTCAAGAATATGCATGTGTGTATGTGTATGGT  109474

Qy       1127 GATT  1130
              ||||
Db     109475 GAAT  109478
```

RESULT 15
AC135226
LOCUS   AC135226   150137 bp   DNA   linear   PLN 05-FEB-2003
DEFINITION Oryza sativa chromosome 3 BAC OSJNBa0037J17 genomic sequence,
            complete sequence.
ACCESSION   AC135226
VERSION   AC135226.4   GI:27805384

```
KEYWORDS    HTG.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 150137)
  AUTHORS   Buell,C.R., Yuan,Q., Ouyang,S., Liu,J., Gansberger,K., Jones,K.M.,
            Overton II,L.L., Tsitrin,T., Kim,M.M., Bera,J.J., Jin,S.S.,
            Fadrosh,D.W., Tallon,L.J., Koo,H., Zismann,V., Hsiao,J., Blunt,S.,
            Vanaken,S.S., Riedmuller,S.B., Utterback,T.T., Feldblyum,T.V.,
            Yang,Q., Haas,B.J., Suh,B.B., Peterson,J.J., Quackenbush,J.,
            White,O., Salzberg,S.L. and Fraser,C.M.
  TITLE     Oryza sativa chromosome 3 BAC OSJNBa0037J17 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 150137)
  AUTHORS   Buell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 150137)
  AUTHORS   Buell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   4  (bases 1 to 150137)
  AUTHORS   Buell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JAN-2003) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   5  (bases 1 to 150137)
  AUTHORS   Buell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-2003) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org
            On Jan 21, 2003 this sequence version replaced gi:24137435.
COMMENT     Address all correspondence to:rice@tigr.org

            BAC clone OSJNBa0037J17 is from Oryza sativa chromosome 3
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of several methods: Gene
            prediction programs including Fgenesh (http://www.softberry.com/),
            genscan and Genscan+ (Chris Burge,
            http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM (Mark Borodovsky,
            http://genemark.biology.gatech.edu/GeneMark/), and GeneSplicer
            (Mihaela Pertea and Steven Salzberg, contact mpertea@tigr.org),
            searches of the complete sequence against a peptide database and
            the plant EST database at TIGR (http://www.tigr.org/tdb/tgi.shtml).
            Annotated genes are named to indicate the level of evidence for
            their annotation. Genes with similarity to other proteins are named
            after the database hits. Genes without significant peptide
            similarity but with EST similarity are named as unknown proteins.
            Genes without protein or EST similarity, that are predicted by more
            than two gene prediction programs over most of their length are
            annotated as hypothetical proteins. Genes encoding tRNAs are
            predicted by tRNAscan-SE (Sean Eddy,
            http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
            identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html).
            This BAC overlaps with rice BACs OSJNBa0034E08 (GB:AC135597) and
            OSJNBb0004TD08 (GB:AC137925).
FEATURES             Location/Qualifiers
     source          1..150137
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="3"
     gene            230..1537
                     /clone="OSJNBa0037J17"
                     /note="japonica cultivar-group"
                     /gene="OSJNBa0037J17.1"
     mRNA            join(<230..577,668..1071,1210..>1537)
                     /gene="OSJNBa0037J17.1"
     CDS             join(230..577,668..1071,1210..1537)
                     /gene="OSJNBa0037J17.1"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO37479.1"
                     /db_xref="GI:28209461"
                     /translation="MEDAEAPRRLLNDTLDLIVGDRERQAYYMLSDREYANTREYSPE
                     LLKKIGMDVEFCATWKAVGWQKFVVDEPSSRLLTLQFICTLKEIEDGISPRFFHKEY
                     TLTWKGLSTLLGFHKPRTNDIHNPTLRLIHKWIAMTWPRGDLRPIREDELIMFAMV
                     RKIKIVPMKYMIRQWLESIKFSAPVECTSLITRIAKGLGVVSDQIAFISATRPCIDET
                     YLVQGHIVKHGINGSLIYFPGCTNEIPLPNAGCMRQVQSQLGMHPDGSKLHATASES
                     QAGRVLVRTGGVRLMTYIRGTTNLPDQVTEEIQYNLMEHIAQTQEWQSVNAQFASIN
                     NMMQQQHDDLQAYFRFQGFNPYQGP"
     repeat_region   1576..1621
                     /rpt_family="tandem"
     gene            complement(4405..5112)
                     /gene="OSJNBa0037J17.2"
                     /note="similar to gypsy type retrotransposon protein orf3
                     GB:BAA84461 GI:5902448 (Oryza sativa (japonica cultivar
                     group)"
     mRNA            complement(<4405..>5112)
                     /gene="OSJNBa0037J17.2"
     CDS             complement(4405..5112)
                     /gene="OSJNBa0037J17.2"
                     /codon_start=1
                     /product="putative retrotransposon protein"
                     /protein_id="AAO37480.1"
                     /db_xref="GI:28209462"
                     /translation="MAPCKPTKAGATGEDPGSLGEGWSCYLGKSLVKESDLADLILTS
                     VLAEGQATCWAKAVVPSPGDRRTVVFAAFFTGLRLPCDDFLPSVLERYEVKLPQLSP
                     LAFPKLAIFAWMCRTCGLAPTAELFAVLFTACTTKDVNTHAGRAEEDSFRLRELYAS
                     AGAFGRLAGACVDGKMGPQLDAEVLRRQPILHRGRQGELAPLSAPSCFHHRQAKCRD
                     RRHTSPGGSFFTTLRGG"
     gene            7834..8349
                     /gene="OSJNBa0037J17.3"
                     /note="predicted by genemarkHMM"
     mRNA            join(<7834..7915,8249..>8349)
                     /gene="OSJNBa0037J17.3"
     CDS             join(7834..7915,8249..8349)
                     /gene="OSJNBa0037J17.3"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO37477.1"
                     /db_xref="GI:28209459"
                     /translation="MAKTIKQRWRRLKDRAGIWQDOMHLGQIVASTTAGAEPGDLRAK
                     ALRANATEKGRGKGC"
     gene            complement(7844..10311)
                     /gene="OSJNBa0037J17.4"
                     /note="predicted by fgenesh"
     mRNA            complement(join(<7844..8124,9092..9316,9474..9567,
                     9622..9897,9973..>10311))
                     /gene="OSJNBa0037J17.4"
     CDS             complement(join(7844..8124,9092..9316,9474..9567,
                     9622..9897,9973..10311))
                     /gene="OSJNBa0037J17.4"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO37481.1"
                     /db_xref="GI:28209463"
                     /translation="MAPCKANPASVMGPDPDRIDDDTTAYLGASLVDEDEVAKLVTSG
                     VLAEKQAFTPGKAVVPKLGDNRTVVFTVFFEAGLRFLCNVLLPEILRLFHVELPQLSP
                     SALFRMAIFDWTVVTPVGMKMAFGSMNFNVRPERSDLWPVNATMSKWDCHWMSKWFY
                     HSIPFEAGSESAKALRWHRRAIALNRKPKVTVDGAMEARFTILRKIFPLFQTWQVEVS
                     PNEEVDGLPKLVMLEGANASCLRGGRPAGQTRCTMANFLPRAKPFKADGDEGTSGNGN
                     SGKSNLPLGSEGPLPPTMTLMTKTLCMTMTAKRKEVPLGFRLYREIYMAALMKGTF
                     AWVYSVIWHGHLMGQLANRPKQLGWPLGPLQSRACYGEAPILQPTHFAQDAPDLARY
                     QRDPSAFAISAL"
     gene            10940..11464
                     /gene="OSJNBa0037J17.5"
                     /note="predicted by fgenesh"
```

```
mRNA            join(<10940. .10998,11443. .>11464)
                /gene="OSJNBa0037J17.5"
CDS             join(10940. .10998,11443. .11464)
                /gene="OSJNBa0037J17.5"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAO37482.1"
                /db_xref="GI:28209464"
                /translation="MAKAEGSRWYLARPNASWAKSYVPMQ"

repeat_region   12527. .12583
                /rpt_family="CT-rich"
gene            complement(12586. .14241)
                /gene="OSJNBa0037J17.6"
mRNA            /note="predicted by genemarkHMM"
                complement(join(<12586. .12666,12994. .13344,13455. .13628,
                13822. .13878,13954. .>14241))
                /gene="OSJNBa0037J17.6"
CDS             complement(join(12586. .12666,12994. .13344,13455. .13628,
                13822. .13878,13954. .14241))
                /gene="OSJNBa0037J17.6"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAO37494.1"
                /db_xref="GI:28209476"
                /translation="MATVTESSWARLAASGGGAGDGRRGRRWAWPAAAEGGAGRRWWE
                VHDDGLKEAKVATTLMWWRDEAERTPVRLTGKRESKLSEEALAAALDPGGKVALGWA
                SSGRWGRSGGRPQRPWIPAARSRSAIDGLLASSRHGVDRTVDGTARSRSAGRRAPEL
                APEFDGGSLRWKKVKAAERRSLWHGDLASDSASVDSDSIDSAYADSSSTSADSASTGD
                SASTMPGVEVAKRAARPGTARARVRHGPFRHGPLGTAHETGRAVPTHVPSRRPKHGPW
                MAGPCRHGMRVARRAVPGHAGYRFRRASGRPFWPGPFGHL"

repeat_region   12719. .12746
                /rpt_family="GC_rich"
repeat_region   13100. .13137
                /rpt_family="tandem"
repeat_region   13111. .13169
                /rpt_family=" (CCGGG)n"
repeat_region   13796. .13852
                /rpt_family="tandem"
repeat_region   14242. .14283
                /rpt_family="G-rich"
repeat_region   14349. .14420
                /rpt_family="tandem"
gene            14705. .18463
                /gene="OSJNBa0037J17.7"
                /note="highly similar to ribosomal protein S5 GB:CAA63650
                GI:5725342 (Spinacia oleracea)"
mRNA            join(<14705. .15435,18214. .>18463)
                /gene="OSJNBa0037J17.7"
CDS             join(14705. .15435,18214. .18463)
```

```
.Query Match        6.4%;      Score 172;  DB 8;  Length 150137;
Best Local Similarity   59.0%;  Pred. No. 3.6e-31;
.Matches  385;  Conservative    0;  Mismatches  260;  Indels   7;  Gaps     5;

Qy    485  TTTTCTTGGCCCATCTTTGGTATTTTCACAAATGTCCCCTACAAATGATAAATCACCAA  544
           |||||||||   ||  ||| |||||||  |||||||  ||||||||||||||||||||
Db   30628  TTTTTGTCGCATGACAGATGTTTTTCTCCTCAAATTACCTACATACACATCTTTCACCAA  30687

Qy    545  AACTCATGGAGCTTGCTAGTTATAAACTCTAATTCTAAGTTTGGT-GTTTATTTGAGTGG  603
           || || || || || ||||   |||  |||||||   |||||||| | ||||||||||
Db   30688  CACTAGTAGAAATTATTAGTAATAATCCCTACCACTAAGTTGGATACCCATATTTTATTGC  30747

Qy    604  ATTTTGTCTGAAAGTTGGTGGTTAGAAATAGGAGTTAAGGACCGCCAACAAGATCCCCCA  663
           |||  |||   ||| |||||||  ||||   |||||  |||  |||| |||||||||
Db   30748  ATTATATGCAGGTATTAACGGTCAAATATTATGATTTAAGGACACCGGTTAGCAACACCCCA  30807

Qy    664  CACTTAGCCCTTTGCTCATCCTCGAGTAAGTTCAAGGACTAAGGTGGAACATCTCCTCA  723
           ||||||   ||  || || ||||||  |  || ||||||  ||  |||||||||||||
Db   30808  CACTTAACCTTTGCTCATCCCGAGTGAAGGCTATGGACCTCTACAACAAAGTTACGACTTC  30867

Qy    724  AATGGTACGATGCCTGCATATAAGTTATTCCAAGCGCTTCACCTATACATGT-GAACTTTGA  782
           |||||  || || |||| ||| |||| |||||  |||||||||||| ||||| |||||||
Db   30868  CTTACTAATAATCACGTAAACAAATTATTCAAAGTCATACCTATACTTGTGGATCTTTTGT  30927
```

```
Qy    783  AGTGTCTACCACGCCATCTTGGGTGGTTGAGAAATGGAACAGATCAGAATCCAGTCATCT  842
           |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   30928  AGTGTCTACCACCGCAATCCTGAGCAATTGAGAAATGGAACAATCTTGACTTAAGTAACCC  30987

Qy    843  TTACCTCCTGCTTAGATAACTTGGGTTTTGTAAGGTTTCAAATTTAAAACATAGTC  902
           |||||||||||||||||   |||||||||||||||||||   |||| || |||
Db   30988  TAACTTTTCATGAAGAACAACTTGGGGTTTTTCAAGTAACACAAAATTATAACTTTGCT  31047

Qy    903  TTGCTCCTCAAATGATTCTCTCATATAGCTCAATGTGTATGGTTTCTCACCAAGGCAATG  962
           || |||||||||||||||||  ||||||||| |  || |||||||||||||||||||
Db   31048  TTACTCCTCGATTGATTCTTC--GATTACTCAAAGTGTATGATTCCTCACAAAGGCATAT  31106

Qy    963  TTTTGCCTCTTTTCAT---CCTACTTCTAATATTTCTTTTGTGGAGCTTAGGGTAGGGAA  1019
           ||| |||||||||||    ||  |||| ||||   |||||||||||||||||||||||
Db   31107  GATTGTGCCTTTTCTTTTTTCCTACCATCTAATAAGGTTTATGTGGAACTTAGGGTATAGAA  31166

Qy   1020  TGAAAAGGAGACATATCTGCATTGCTTGCGATATGTTACTAAGTCAAAAACCAAATCTGAGGAGA  1079
           ||||| ||||||||||||||||||||||||| ||||||| |||| |||||||||||
Db   31167  AAGAGTAGAGCATTGCTTGCATCCACATATTGCAAAGTCAAATAATGGATCCACCGGAAA  31226

Qy   1080  AGCAAGTCATACAATCTGATCAAG-ATGTGCAAGTGTGTGGATATGTGGATT  1130
           |||| ||||    || || |||||| |||||||||||||||||| ||||||
Db   31227  TACAAGACATACTACTTTCTGGTCAAGAAATGTGAATGTGTATGTGTACGGTGAAT  31278
```

Search completed: May 9, 2005, 01:51:44
Job time : 11595 secs

THIS PAGE LEFT BLANK

```
                    GenCore version 5.1.6
            Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:    May  8, 2005, 21:22:28 ; Search time 1413 Seconds
                             (without alignments)
                             11269.713 Million cell updates/sec

Title:             US-10-751-550-1
Perfect score:     2690
Sequence:          1 tctagataatacgactcact.........ccataactagtaccaccatgg 2690

Scoring table:     IDENTITY_NUC
                   Gapop 10.0 , Gapext 1.0

Searched:          439206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters:      8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       N Geneseq_16Dec04:*
            1: geneseqn1980s:*
            2: geneseqn1990s:*
            3: geneseqn2000s:*
            4: geneseqn2001as:*
            5: geneseqn2001bs:*
            6: geneseqn2002as:*
            7: geneseqn2002bs:*
            8: geneseqn2003as:*
            9: geneseqn2003cs:*
           10: geneseqn2003ds:*
           11: geneseqn2004as:*
           12: geneseqn2004bs:*
           13: geneseqn2004bs:*

        Pred. No. is the number of results predicted by chance to have a
        score greater than or equal to the score of the result being printed,
        and is derived by analysis of the total score distribution.

                              SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2690	100.0	2690	13	ADR21934	Adr21934 Sugarcane
2	1001.2	37.2	1047	4	AAS01032	Aas01032 Sugarcane
3	151.4	5.6	2000	12	ADJ41549	Adj41549 Plant cDN
4	149.2	5.5	4797	12	ADJ40495	Adj40495 Plant cDN
5	114.4	4.3	2000	12	ADJ41110	Adj41110 Plant cDN
6	55.4	2.1	2000	8	ADA71938	Ada71938 Rice gene
7	53.4	2.0	51259	2	AAX83007	Aax83007 Partial m
8	52.2	1.9	72352	12	ADQ97067	Adq97067 Mouse can
9	51.8	1.9	3718	2	AAV72244	Aav72244 G. max SB
10	51	1.9	1150	8	ABZ57867	Abz57867 Porcine e
11	51	1.9	2860	8	ABZ57861	Abz57861 Porcine e
12	50.8	1.9	1976	8	ABZ57865	Abz57865 Porcine e
13	50	1.9	267	4	AAK45604	Aak45604 Human bon
14	50	1.9	267	4	AAK19599	Aak19599 Human bra
15	50	1.9	267	4	ABS45294	Abs45294 Human liv
16	50	1.9	267	6	ABS19876	Abs19876 Human gen
17	50	1.9	474	4	AAK32380	Aak32380 Human bon
18	50	1.9	474	4	AAK06682	Aak06682 Human bra
19	50	1.9	474	4	ABS32088	Abs32088 Human liv
20	50	1.9	474	6	ABS07163	Abs07163 Human gen

21	50	1.9	943	12	ADM72367	Adm72367 Soybean A
22	50	1.9	2385	3	AAC83331	Aac83331 PART-1 pr
23	50	1.9	3809	9	ABK95303	Abk95303 Human pro
24	49.6	1.8	2623	4	AAD17484	Aad17484 Mouse glu
25	49.6	1.8	3815	8	ABS57008	Abs57008 Genomic D
26	49.2	1.8	795	2	AAV55830	Aav55830 FLGA inse
27	49.2	1.8	799	2	AAV55831	Aav55831 Nucleotid
28	49.2	1.8	1925	2	AAX90924	Aax90924 Epstein B
29	49.2	1.8	1926	2	AAA50254	Aaa50254 Epstein B
30	49.2	1.8	1926	4	AAF82902	Aaf82902 EBV tethe
31	49.2	1.8	1926	10	ADK65580	Adk65580 Human her
32	49.2	1.8	2580	3	AAA75454	Aaa75454 Nucleotid
33	49.2	1.8	2580	6	AAI64275	Aai64275 Epstein-B
34	49.2	1.8	5452	2	AAX90923	Aax90923 Anti-sens
35	49.2	1.8	8705	2	AAZ23778	Aaz23778 Vector pS
36	49.2	1.8	8705	12	ADM10659	Adm10659 Expressio
37	49.2	1.8	9482	12	ADP64415	Adp64415 Vector pC
38	49.2	1.8	9600	2	AAV21683	Aav21683 Vector pl
39	49.2	1.8	10285	6	ABS71027	Abs71027 pCEP-Xa-F
40	49.2	1.8	10285	6	ABS66453	Abs66453 Plasmid p
41	49.2	1.8	10330	12	ADL67154	Adl67154 Plasmid p
42	49.2	1.8	10380	2	AAZ22248	Aaz22248 Nucleotid
43	49.2	1.8	10477	12	ADL67152	Adl67152 Plasmid p
44	49.2	1.8	10516	12	ADL67150	Adl67150 Plasmid p
45	49.2	1.8	10561	12	ADL67148	Adl67148 Plasmid p

```
                              ALIGNMENTS

RESULT 1
ADR21934
ID   ADR21934 standard; DNA; 2690 BP.
XX
AC   ADR21934;
XX
DT   21-OCT-2004  (first entry)
XX
DE   Sugarcane jasmonate-induced protein (JAS) promoter DNA SeqID 1.
XX
KW   jasmonate-induced protein; JAS; promoter; stem-specific;
KW   defence-inducible; carbon metabolism; insecticidal; pest tolerance;
KW   plant; ds.
XX
OS   Saccharum.
XX
FH   Key            Location/Qualifiers
FT   CAAT_signal    2428..2431
FT                  /*tag= a
FT   TATA_signal    2604..2610
FT                  /*tag= b
FT   misc_feature   2687..2689
FT                  /*tag= c
FT                  /note= "Start codon (AUG) of JAS"
XX
PN   WO2004062366-A2.
XX
PD   29-JUL-2004.
XX
PF   05-JAN-2004; 2004WO-US000115.
XX
PR   03-JAN-2003; 2003US-0437974P.
XX
PA   (TEXA ) UNIV TEXAS A & M SYSTEM.
XX
PI   Damaj MB, Mirkov ET, Reddy AS, Thomas TL, Rathore KS, Emani C;
PI   Kumpatla SP;
XX
DR   WPI; 2004-544019/52.
XX
PT   New isolated nucleic acid comprises jasmonate-induced protein (JAS)
PT   promoter and an exogenous nucleic acid, useful as a promoter for altering
PT   carbon metabolism in a plant cell or for driving expression of
```

PT   insecticidal proteins in sugarcane.
XX
PS   Claim 1; SEQ ID NO 1; 58pp; English.
XX
CC
CC   This invention relates to a novel nucleic acid molecule that comprises a
CC   jasmonate-induced protein (JAS) promoter and an exogenous nucleic acid.
CC   Specifically, it refers to a stem-specific promoter active in plants that
CC   exhibits enhanced specificity in regulating gene expression in stem
CC   tissues and in response to induction by external stimuli such as plant
CC   defence-inducing agents. The present invention describes a bacterial cell
CC   comprising an expression vector that can be used to transform a monocot
CC   plant such as sugarcane, sorghum, rice, maize or hybrids thereof, in
CC   order to upregulate localised expression. In particular, this nucleic
CC   acid is useful as a promoter for altering carbon metabolism in the
CC   sucrose accumulating tissues, and for driving expression of insecticidal
CC   proteins in sugarcane. These promoters may also be applied to the
CC   development of improved pest and disease tolerant rice plants. This
CC   polynucleotide sequence is the JAS promoter DNA sequence of the
CC   invention.
XX
SQ   Sequence 2690 BP; 732 A; 608 C; 586 G; 764 T; 0 U; 0 Other;

Query Match   100.0%; Score 2690; DB 13; Length 2690;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2690; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 TCTAGATAATACGACTCACTATAGGGCACGCGTCGGTCGACGGCCCGGGTCGGTCTGCGAC   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TCTAGATAATACGACTCACTATAGGGCACGCGTCGGTCGACGGCCCGGGTCGGTCTGCGAC   60

Qy     61 AGCTAGAGGGCGCCACCGGGTCCTAGCTTCCTCCCAACTTCTCGTCGGAGATCCCTTCAGGG  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AGCTAGAGGGCGCCACCGGGTCCTAGCTTCCTCCCAACTTCTCGTCGGAGATCCCTTCAGGG  120

Qy    121 ATGCCCAATGCCACCGCCCCTAAGTCAACCTGCCGGAGCTGGAGCTTCGCCAGGTCAGA     180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ATGCCCAATGCCACCGCCCCTAAGTCAACCTGCCGGAGCTGGAGCTTCGCCAGGTCAGA     180

Qy    181 GCTGCGGCAGCACCCTGGTAGACCGGCTCCGATGATCCTGATGACCCGGTCGCTCCATGAAG  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GCTGCGGCAGCACCCTGGTAGACCGGCTCCGATGATCCTGATGACCCGGTCGCTCCATGAAG  240

Qy    241 AAGTGCATTCGCCCAACCAAGTCGAGTGGGTCGCCTGGAGGGGGCGGGGAAGCAAAACGT    300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AAGTGCATTCGCCCAACCAAGTCGAGTGGGTCGCCTGGAGGGGGCGGGGAAGCAAAACGT    300

Qy    301 TGCATGCACCTAGGCGCCCTGGCAGGAGCTCCTGTAGTATCACCTGCGTCGCCTCCAGCT    360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TGCATGCACCTAGGCGCCCTGGCAGGAGCTCCTGTAGTATCACCTGCGTCGCCTCCAGCT    360

Qy    361 CATGCTCCGCAAGCCTCCAGGCGCGGCCCGGCAGTGCTCCAACACTTCGCCTCCTCCTACA   420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CATGCTCCGCAAGCCTCCAGGCGCGGCCCGGCAGTGCTCCAACACTTCGCCTCCTCCTACA   420

Qy    421 GCTCCTTCCACATGCAGTCGTGCTCCGCACGCACCTTCCACCTTTTTTACTCTTTTTCTT   480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GCTCCTTCCACATGCAGTCGTGCTCCGCACGCACCTTCCACCTTTTTTACTCTTTTTCTT   480

Qy    481 TCTCTTTTCTTGGCCCATCTTTGGTATTTTCACAAATGTCCCCTACAAATGATAAATCA    540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TCTCTTTTCTTGGCCCATCTTTGGTATTTTCACAAATGTCCCCTACAAATGATAAATCA    540

Qy    541 CCAAAACTCATGGAGCTTGCTAGTTATAAAACTCTAAGTTTGTGTTTTATTTGAG       600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CCAAAACTCATGGAGCTTGCTAGTTATAAAACTCTAAGTTTGTGTTTTATTTGAG       600

Qy    601 TGGATTTGTGTGAAAGTTGGTGGTTAGAAATAGGAGTTAAGGACCGGCCAACAAGATCCC   660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TGGATTTGTGTGAAAGTTGGTGGTTAGAAATAGGAGTTAAGGACCGGCCAACAAGATCCC   660

Qy    661 CCACACTTAGCCCTTTGCTCATCCTCGAGTAAGTTCAAGGACTAAGGTGGAACATCTCC   720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CCACACTTAGCCCTTTGCTCATCCTCGAGTAAGTTCAAGGACTAAGGTGGAACATCTCC   720

Qy    721 TCAAATGGTACGATGCCTGCATATAAGTATTATTCCAAGCCTCACCTATACATGTGAACTTT  780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 TCAAATGGTACGATGCCTGCATATAAGTATTATTCCAAGCCTCACCTATACATGTGAACTTT  780

Qy    781 GAAGTGTCTACCACGCCATCTTGGGTGGTGGTGAGAAATGGAACAGATCAGAATCCAGTCAT  840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GAAGTGTCTACCACGCCATCTTGGGTGGTGGTGAGAAATGGAACAGATCAGAATCCAGTCAT  840

Qy    841 CTTTACCTCTTGTTGTAGATAACTTGGGTTTTGTAAGGTTTCAAATTTAAAACATAG      900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CTTTACCTCTTGTTGTAGATAACTTGGGTTTTGTAAGGTTTCAAATTTAAAACATAG      900

Qy    901 TCTTGCTCCCAAATGATTCTCTCATATAGTCAATGTGATGGTTCTCACCAAGGCAA      960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TCTTGCTCCCAAATGATTCTCTCATATAGTCAATGTGATGGTTCTCACCAAGGCAA      960

Qy    961 TGTTTTGCCTCTTTTCATCCTACTTCTAATATTTCTTTTGTGGAGCTTAGGGTAGGGAAT  1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TGTTTTGCCTCTTTTCATCCTACTTCTAATATTTCTTTTGTGGAGCTTAGGGTAGGGAAT  1020

Qy   1021 GAAAAGGAAGCATACTGCATTGCATATGTTACTAAGTCAAAACCAAATCTGACGAGAA    1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GAAAAGGAAGCATACTGCATTGCATATGTTACTAAGTCAAAACCAAATCTGACGAGAA    1080

Qy   1081 GCAAGTCATCACAATCTGATCAAGATGTGCAAGTGTGTGGATATGTGGATTAAGATAACTC  1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 GCAAGTCATCACAATCTGATCAAGATGTGCAAGTGTGTGGATATGTGGATTAAGATAACTC  1140

Qy   1141 CTGTTTATTCATGCTCTCCTCCTAATAAACTTTAGAGGGCATGGCAATCTTTGCATGGG   1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CTGTTTATTCATGCTCTCCTCCTAATAAACTTTAGAGGGCATGGCAATCTTTGCATGGG   1200

Qy   1201 CCTTCATGAGCTCATCGTATGTCTAAGCATGGAGCTCATCATTTATATAGACATGGTGAT   1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CCTTCATGAGCTCATCGTATGTCTAAGCATGGAGCTCATCATTTATATAGACATGGTGAT   1260

Qy   1261 ACCAAAATTACTCCTTTTGAGCATGTTTATATTTACCTGTTGAGGTA               1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 ACCAAAATTACTCCTTTTGAGCATGTTTATATTTACCTGTTGAGGTA               1320

Qy   1321 AATCTGAACGCTAATAATCGGCTAAGCAAATAATTATCACCTGTTGATTCTAACAAT     1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 AATCTGAACGCTAATAATCGGCTAAGCAAATAATTATCACCTGTTGATTCTAACAAT     1380

Qy   1381 TTGATGATGACAATATTGATGAGTGACTGACAAATGATTGAAGGCTTTAAAGGAGATT    1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TTGATGATGACAATATTGATGAGTGACTGACAAATGATTGAAGGCTTTAAAGGAGATT    1440

Qy   1441 GAGAAGGATAAATCTACAATAAAAATGTAAAGATAAAAATGTAAAGAACAACAAGTGAATCTCG 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GAGAAGGATAAATCTACAATAAAAATGTAAAGATAAAAATGTAAAGAACAACAAGTGAATCTCG 1500

Qy   1501 TGTGGAAGACTATTTGCCTCTTGGGGGTAAAAGACAACAAGTTAGTAAGTTAGTGGCCTCAA  1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TGTGGAAGACTATTTGCCTCTTGGGGGTAAAAGACAACAAGTTAGTAAGTTAGTGGCCTCAA  1560

Qy   1561 AATTGGGAGGGCCCATGCAAGATGTTAAAGTAATTGTTTGGATTGACGGAGGCATTTC    1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 AATTGGGAGGGCCCATGCAAGATGTTAAAGTAATTGTTTGGATTGACGGAGGCATTTC    1620

Qy   1621 AAGGTGATCATCTACCTGAGACGTCTCAATGGGAGGTGCTCGAAGACATATTACCCATGTG  1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 AAGGTGATCATCTACCTGAGACGTCTCAATGGGAGGTGCTCGAAGACATATTACCCATGTG  1680

Qy   1681 TATGGCAAGATGTTAGCTAGTAACTGACGATAGTGTAAACGATCTCCAATGGGGCAAG    1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 TATGGCAAGATGTTAGCTAGTAACTGACGATAGTGTAAACGATCTCCAATGGGGCAAG    1740

Qy   1741 ACATATTACCTAAGGCCAGCCAGGCTGGTTTTTCGAGTTCGAGTATAGAGATTCTCGT    1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 ACATATTACCTAAGGCCAGCCAGGCTGGTTTTTCGAGTTCGAGTATAGAGATTCTCGT    1800
```

```
Qy  1801  GCGAGTTGTAAACGATCTCCAATGGGCAAGACATCCTACCCTATATAGTGAAGGGGC  1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1801  GCGAGTTGTAAACGATCTCCAATGGGCAAGACATCCTACCCTATATAGTGAAGGGGC  1860

Qy  1861  AGTAGCTGATTGAGAATCAATCAATCAAGCACAATAATÁTAATTTATTAATTTTTTATTCAAA  1920
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db  1861  AGTAGCTGATTGAGAATCAATCAATCAAGCACAATAATATAATTTATTAATTTTTTATTCAAA  1920

Qy  1921  CCCAATTTTTTTCCTTTTCCAACCCTAATTATATAGTGTTTCCTTTTGCCTCTAGGACAAATT  1980
          ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db  1921  CCCAATTTTTTTCCTTTTCCAACCCTAATTATATAGTTTTCCTTTTGCCTCTAGGACAAATT  1980

Qy  1981  GACGTGTTCCGGGTATCCTGCTGAATTAAGAACAACACCCTAGGTGCACCTGTCCCGATAGA  2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1981  GACGTGTTCCGGGTATCCTGCTGAATTAAGAACAACACCCTAGGTGCACCTGTCCCGATAGA  2040

Qy  2041  GTCCCACCTGGGTAGGCATTCATAGGGATTCGTGTATTTCCTGCAAAAAAGCGATTAAGC  2100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041  GTCCCACCTGGGTAGGCATTCATAGGGATTCGTGTATTTCCTGCAAAAAAGCGATTAAGC  2100

Qy  2101  TGGCTTCTAAAACTGGCTAGGCCGGATTCGTGTGCCTTCACTACCAGGTGATTTTCATGT  2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101  TGGCTTCTAAAACTGGCTAGGCCGGATTCGTGTGCCTTCACTACCAGGTGATTTTCATGT  2160

Qy  2161  GATCCGTGCATTCTAGCACTTGCTATGTAACCCAAACTTAAGTCGACAACTATAAAATAT  2220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161  GATCCGTGCATTCTAGCACTTGCTATGTAACCCAAACTTAAGTCGACAACTATAAAATAT  2220

Qy  2221  GCTACTTGCAGGATGTTATCACGACACAACTCCTAATCTACGGAAGCCTAAGTTTAGTTT  2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221  GCTACTTGCAGGATGTTATCACGACACAACTCCTAATCTACGGAAGCCTAAGTTTAGTTT  2280

Qy  2281  TGCTCGGAGGACAAGCAATTGTGGCCAGTCACTATAGCTACGTCAGAGGGGTAGTGGGAGCA  2340
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281  TGCTCGGAGGACAAGCAATTGTGGCCAGTCACTATAGCTACGTCAGAGGGGTAGTGGGAGCA  2340

Qy  2341  GTTGCGTCGTTGGATTGAAAACAGGTGGATCGTATCAGATATTATGCATTCACATGGACA  2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341  GTTGCGTCGTTGGATTGAAAACAGGTGGATCGTATCAGATATTATGCATTCACATGGACA  2400

Qy  2401  GTAAATGTGGTACAGTAACTTCGCAAACAATAAAATCTGTCACAATTTATTAGTGCACTC  2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401  GTAAATGTGTACAGTAACTTCGCAAACAATAAAATCTGTCACAATTTATTAGTGCACTC  2460

Qy  2461  CTCTGACGTAAATGCTTCTACGTCAGAGGATTGATTCCGAGGGCCGCTGCACCCATCAC  2520
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461  CTCTGACGTAAATGCTTCTACGTCAGAGGATTGATTCCGAGGGCCGCTGCACCCATCAC  2520

Qy  2521  TAATGACGGTCTTTACCCATCATCATGGACCATTGTTCACATCCATGTCATCACTGTCGT  2580
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521  TAATGACGGTCTTTACCCATCATCATGGACCATTGTTCACATCCATGTCATCACTGTCGT  2580

Qy  2581  CCTGTCCATGCACTGCAGCCCTTCTATAAATACTGGCATCCTCCCCCGTTCACAGATCAC  2640
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2581  CCTGTCCATGCACTGCAGCCCTTCTATAAATACTGGCATCCTCCCCCGTTCACAGATCAC  2640

Qy  2641  ACAACACAAGCAAGAAATAAACGTAGCTGCCATAACTAGTACCACCATGG  2690
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2641  ACAACACAAGCAAGAAATAAACGTAGCTGCCATAACTAGTACCACCATGG  2690
```

```
KW   beta-glucuronidase; GUS; starch biosynthesis; fatty acid biosynthesis;
KW   ADP-glucose pyrophosphorylase; sucrose metabolism; 67pro; ds.
OS   Saccharum sp.
PN   WO200118211-A1.
PD   15-MAR-2001.
PF   01-SEP-2000; 2000WO-AU001033.
PR   02-SEP-1999; 99AU-00002625.
PA   (UYQU ) UNIV QUEENSLAND.
PI   Potier B, Birch RG;
DR   WPI; 2001-218560/22.
PT   New sugarcane plant promoters for directing expression of heterologous
PT   nucleic acids in a constitutive or tissue-specific manner in
PT   monocotyledonous plants.
PS   Claim 10; Fig 18; 107pp; English.

CC   The present sequence for sugarcane stem-specific gene promoter 67pro is 1
CC   of 4 promoter regions of specific transcribed DNA sequences (AAS01032-
CC   AAS01035). Also described are 11.promoter regions of a transcribable DNA
CC   sequence isolated from various sugarcane cDNA clones (AAS01021-AAS01031).
CC   The nucleic acids are useful for producing transgenic plants, having an
CC   altered phenotype and for driving expression of a foreign or endogenous
CC   DNA sequence, which encode agronomic properties including insecticide,
CC   herbicide, disease resistance, stress tolerance and improved food
CC   content, or increased yields. The foreign or endogenous DNA sequence may
CC   comprise a region transcribed into an antisense RNA or ribozyme that
CC   modulates the expression of a corresponding target gene, or it may encode
CC   beta-glucuronidase (GUS), luciferase, neomycin phosphotransferase, a
CC   product conferring herbicide tolerance, a product affecting starch
CC   biosynthesis or modification, ADP-glucose pyrophosphorylase, a product
CC   involved in fatty acid biosynthesis, a product conferring insect
CC   resistance, a product altering sucrose metabolism or a gene encoding
CC   valuable pharmaceuticals, e.g. antibiotics, secondary metabolites or
CC   vaccines. The promoters are capable of directing high level expression in
CC   many or all cells of a plant, preferentially in stem or meristem tissue
CC   of monocotyledonous plants.
SQ   Sequence 1047 BP; 297 A; 240 C; 220 G; 290 T; 0 U; 0 Other;

Query Match            37.2%;    Score 1001.2;    DB 4;    Length 1047;
Best Local Similarity  98.6%;    Pred. No. 5.7e-263;
Matches 1031;    Conservative  0;    Mismatches 13;    Indels 2;    Gaps  2;
```

```
Qy  1639  GAGCTCTCAATGGGAGGTGCTCGAAGACATATTACCCATGTGTATGGCAAGATGTTTAGC  1698
          |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db     1  GAGCTCTCAATGGGAGGTGCTCGAAGACATATTACCCAAGTGTATGGCAAGATGTTTAGC    60

Qy  1699  TAGTAACTGACTGATGAGTGTAAACGAAGATCATCCCAATGGGCAAGACATATTACCTAAGGCCA  1758
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  TAGTAACTGACTGATGAGTGTAAACGAAGATCATCCCAATGGGCAAGACATATTACCTAAGGCCA   120

Qy  1759  GGGCTGGTTTTTGCAAGTTCGAGTAGGATATAGAGATTCTCGTGCGAGTTGTAAACGATCT  1818
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  GGGCTGGTTTTTGCAAGTTCGAGTAGGATATAGAGATTCTCGTGCGAGTTGTAAACGATCT   180

Qy  1819  CCAATGGGGCAAGACATCCTACCCTATATATAGTGAAGGGGCAGTAGCTGATTGAGAATC  1878
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  CCAATGGGGCAAGACATCCTACCCTATATATAGTGAAGGGGCAGTAGCTGATTGAGAATC   240

Qy  1879  AATCAATCAAGCACACAATATAATTTATTAATTTTTTATTCAAACCCAATTTTTTTCCTTTT  1938
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  AATCAATCAAGCACACAATATAATTTATTAATTTTTTATTCAAACCCAATTTTTTTCCTTTT   300

Qy  1939  CCAACCCTAATTATATAGTTTTCCTTTTGCCTCTAGGACAAATTGACGTGTTCCGGGTATCC  1998
```

```
RESULT 2
AAS01032
ID   AAS01032 standard; DNA; 1047 BP.
XX
AC   AAS01032;
XX
DT   29-MAY-2001 (first entry)
XX
DE   Sugarcane stem-specific gene promoter sequence 67pro.
XX
KW   Sugarcane promoter region; monocotyledonous plant; stem tissue;
KW   insecticide; herbicide; disease resistance; improved food content;
```

```
Db    301 |||||||||||||||||||||||||||||||| ||||||||||||
          CCAACCCTAATTATAGTTCTCCTTTGCCTCTAGACAAATTGACGTGTTCCGGGTATCC 360

Qy   1999 TGCTGAATTAAGAACAACCCTAGGTGCACCCTGTCCCGATAGAGTCCACCTGGGTAGGCA 2058
          |||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db    361 TGCTGAATCAAGAACAACACCCCTAGGTGCACCCTGTCCCGATAGAGTCCACCTGGGTAGGCA 420

Qy   2059 TTCATGGGGATTCGTGTATTCCTGCAAAAAAGCGATTAAGCTGGCTTCAAAAACTGGCT 2118
          ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db    421 TTCATGGGGATTCGGGTATTCCTGCAAAAAAGCGATTAAGCTGGCTTCAAAAACTGGCT 480

Qy   2119 AGGCCGGATTCGTGTGGCCTTCACTACCAGGTGATTTTCATGTGATCCGTGCATTCTAGCA 2178
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AGGCCGGATTCGTGTGGCCTTCACTACCAGGTGATTTTCATGTGATCCGTGCATTCTAGCA 540

Qy   2179 CTTTGCTATGTAACCCAAACTTAAGTCGACAACTATAAATATGCTACTTGCAGGATGTTA 2238
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CTTTGCTATGTAACCCAAACTGAAGTCGACAACTATAAATATGCTACTTGCAGGATGTTA 600

Qy   2239 TCACGACACAACTCCTAATCTACGGAAGCCTAAGTTAGTTTGCTCGAGACAAGCAAT 2298
          ||||||||||||||||||||||||||||| ||||||| |||||||| ||||||||||||
Db    601 TCACGACACAACTCCCAGTCTAC-GAAGCCTAAGTTAGTTTGCTCGAGACAAGCAAT 659

Qy   2299 TGTGGCCAGTCACTATAGCTACGTCAGAGGGGTAGTGGGAGCAGTTGCGTCGTTGGATTGA 2358
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    660 TGTGGCCAGTCACTATAGCTACGTCAGAGGGGTAGTGGGAGCAGTTGCGTCGTTGGATTGA 719

Qy   2359 AAACAGGTGATCGTATCAGATATTATGCATTCACATGGACAGTAAATGTGGTACAGTAA 2418
          ||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db    720 AAACAGGTGATCGTATCAGATATTATGCATTCACATGGACAGTAAATGTGGTACAGT-A 778

Qy   2419 CTTGCAACAATAAAATCTGTCACAATTTATTAGTGCACTCCTCGACGTAAATGCTTC 2478
          |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db    779 CTTCGCAAACAATAAAATCTGTCACAATTTATTAGTGCACTCCTCGACGTAAATGCTTC 838

Qy   2479 TACGTCAGAGGATTTGATTCGGAGGGCCGCGGAGGGCGCCACCATCACTAATGACGGTCTTTACCC 2538
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    839 TACGTCAGAGGATTTGAGTCGGAGGGCTGCTGACACATCACTAATGACGGTCTTTACCC 898

Qy   2539 ATCATCATGGACCATTGTTCACATCCATCCATCGTGTCCTGTCCATGCACTGCAG 2598
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    899 ATCATCATGGACCATTGTTCACATCCATCCATCGTGTCCTGTCCATGCACTGCAG 958

Qy   2599 CCCTCTATAAATACTGGCATCCCTCCCCCGTTCACAGATCACACAACACAAGCAAGAAAT 2658
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    959 CCCTCTATAAATACTGGCACCCCCTCCCCCGTTCACAGATCACCACCACAAGCAAGAAAT 1018

Qy   2659 AAACGGTAGCTGCCATAACTAGTACA 2684
          ||||||||||||||||||||||||||
Db   1019 AAACGGTAGCTGCCATAACTAGTACA 1044
```

RESULT 3
AD741549
ID   AD741549   standard; cDNA; 2000 BP.
XX
AC   AD741549;
XX
DT   06-MAY-2004   (first entry)
XX
DE   Plant cDNA #2549.
XX
KW   Plant; gene; ss; transcription; plant genome augmentation; cereal;
KW   soybean; alfalfa; sunflower; canola; cotton; peanut; tobacco; sugar beet;
KW   maize; barley; sorghum; rice; wheat; crop plant; insecticide resistance;
KW   stress tolerance; salt tolerance; cold tolerance; drought tolerance;
KW   plant nutrition; apical dominance; dwarfism; early flowering; antiviral;
KW   antifungal.
XX
OS   Eukaryota.
XX
PN   US2004016025-A1.
XX

PD   22-JAN-2004.
XX
PF   26-SEP-2002; 2002US-00260238.
XX
PR   26-SEP-2001; 2001US-0325277P.
PR   26-SEP-2001; 2001US-0325448P.
PR   04-APR-2002; 2002US-0370620P.
XX
PA   (BUDW/)   BUDWORTH P.
PA   (MOUG/)   MOUGHAMER T.
PA   (BRIG/)   BRIGGS S P.
PA   (COOP/)   COOPER B.
PA   (GLAZ/)   GLAZEBROOK J.
PA   (GOFF/)   GOFF S A.
PA   (KATA/)   KATAGIRI F.
PA   (KREP/)   KREPS J.
PA   (PROV/)   PROVART N.
PA   (RICK/)   RICKE D.
PA   (ZHUT/)   ZHU T.
XX
PI   Budworth P, Moughamer T, Briggs SP, Cooper B, Glazebrook J;
PI   Goff SA, Katagiri F, Kreps J, Provart N, Ricke D, Zhu T;
XX
DR   WPI; 2004-190374/18.
XX

PT   New rice promoter, useful for manipulating crop plants to alter or
PT   improve phenotypic characteristics, e.g. produce large quantities of oil
PT   or proteins, resistance to insecticides, virus or fungi, stress tolerance
PT   or high nutritional value.
XX
PS   Claim 1; SEQ ID NO 2549; 230pp; English.
XX
CC   The invention relates to plant nucleotide sequences that direct seed-,
CC   leaf- and/or stem-, panicle-, root- or pollen-specific or -preferential
CC   or constitutive transcription of an operatively linked nucleic acid
CC   segment. The invention also relates to a method for augmenting a plant
CC   genome and a method of identifying a gene, where its expression is
CC   altered in the seed, leaf, stem, panicle, pollen, root or is constitutive
CC   in a plant cell. The plant is a cereal, e.g. soybean, alfalfa, sunflower,
CC   canola, cotton, peanut, tobacco or sugar beet, preferably maize, barley,
CC   sorghum, rice or wheat. The polynucleotides and the polypeptides they
CC   encode are useful for manipulating crop plants to alter or improve
CC   phenotypic characteristics, to produce large quantities of oil or
CC   proteins, to incur resistance to insecticides, viruses or fungi, and to
CC   incur stress tolerance (e.g. salt, cold or drought) to ensure the plants
CC   have a high nutritional value.with reduced apical dominance or dwarfism,
CC   early flowering or altered metabolic pathways. This sequence represents a
CC   plant nucleic acid of the invention. Note: The sequence data for this
CC   patent did not form part of the printed specification but was obtained in
CC   electronic format directly from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 2000 BP; 537 A; 374 C; 551 G; 536 T; 0 U; 2 Other;

Query Match   5.6%; Score 151.4; DB 12; Length 2000;
Best Local Similarity 61.6%; Pred. No. 3.3e-30;
Matches 305; Conservative 0; Mismatches 166; Indels 24; Gaps 3;

```
Qy   1232 GAGCTCATCATTATATAAGACATGTGATACCAAAATTACTCCTTTGAGCATGTTTATA 1291
          ||| ||||| ||||| || |||||||||| |||||||||||||||||||||||||||
Db     86 GGGCCGCATAGGATATCTAAACATGTGCCACTAAGTTACTCCTTTGAGTTGGTGTATG 145

Qy   1292 TTTAGGAGACGTTTTACCTGTTGAGGTAAATCTGAACCGCTAATAAATCGGCTAAGCAAA 1351
          |||  || | ||||  |||||||||||| ||||||||||||||||||||||||||||||
Db    146 GTCAAGAAGCCGTTTTACCGTGGGAGGTAAATCTTGGTCCTCTTCGTTATATCAGCAAG 205

Qy   1352 ATAATTATTCACCTGTTGATTCTAACAATTTGATGATGATGACAATATTGATGAGGTGACTG 1411
          ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db    206 ATGATTTGTCAAGTGAAGATTACAAGATGTTGATGGGACAAATCTTGATGAAGTCATCG 265

Qy   1412 ACAAATGATTGAAGGCTTAAAGGAGGATTGAGAAGGATAA------------ATCT 1455
          ||||||||||||||||||||| || |||||| |||||||||            ||||
Db    266 ACAAGCGCTTCGAAGGCATTGGAAGGCATTGGAAGAGGATAGAAGGAAGAGAGGGGTCGCCAAGGCGT 325
```

```
CC   The invention relates to plant nucleotide sequences that direct seed-,
CC   leaf- and/or stem-, panicle-, root- or pollen-specific or -preferential
CC   or constitutive transcription of an operatively linked nucleic acid
CC   segment. The invention also relates to a method for augmenting a plant
CC   genome and a method of identifying a gene, where its expression is
CC   altered in the seed, leaf, stem, panicle, pollen, root or is constitutive
CC   in a plant cell. The plant is a cereal, e.g. soybean, alfalfa, sunflower,
CC   canola, cotton, peanut, tobacco or sugar beet, preferably maize, barley,
CC   sorghum, rice or wheat. The polynucleotides and the polypeptides they
CC   encode are useful for manipulating crop plants to alter or improve
CC   phenotypic characteristics, to produce large quantities of oil or
CC   proteins, to incur resistance to insecticides, viruses or fungi, and to
CC   incur stress tolerance (e.g. salt, cold or drought) to ensure the plants
CC   have a high nutritional value with reduced apical dominance or dwarfism,
CC   early flowering or altered metabolic pathways. This sequence represents a
CC   plant nucleic acid of the invention. Note: The sequence data for this
CC   patent did not form part of the printed specification but was obtained in
CC   electronic format directly from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 4797 BP; 1387 A; 934 C; 1260 G; 1209 T; 0 U; 7 Other;

Query Match             5.5%;  Score 149.2;  DB 12;  Length 4797;
Best Local Similarity   61.7%;  Pred. No. 2.2e-29;
Matches  301;  Conservative    0;  Mismatches  163;  Indels   24;  Gaps    3;

Qy      1232 GAGCCTCCATCATTTATATAAGCATGGTGATACCAAAATTACTCCTTTGAGCATGTTTATA 1291
             ||||||||||||||| ||||||||||| |||||||  |||| ||||||| ||||||| ||
Db      4310 GGGCACATAGGATATCTAAACATGTGCCACTAAAGTCACTCCTTTGAGTTGGTTTATG 4369

Qy      1292 TTTAGGAGGACGTTTTACCTGTTGAGGTAAATCTGAACGCTAATAAATCGGCTAAGCAAA 1351
             |||| ||||||||||||||| |||||||||||||  ||| |||| | ||||  |||||||
Db      4370 GTCAAGAAGCCGTTTTACCGTGGAGGTAAATCTTGATTCCTTCGTTATATCAAGCAAG 4429

Qy      1352 ATAATTTATCACCTGTTGATTCTAACAATTGATGATGACAATATTGATGAGGTGACTG 1411
             || |||| |||| ||||||||||| |||| ||||| ||| |||| ||||||| |||||
Db      4430 ATGATTTGTCAAGTGAAGATTACAGACGTTGATGGGAGCAACCTTGATGAAGTCATTG 4489

Qy      1412 ACAAATGATTGAAGGCTTTAAAGGATGGATTGAGAACGGATAA-----------ATCT 1455
             ||||  ||||||||| ||| || |||  || ||| |||||| ||||
Db      4490 ACAAGCGCTTGAAGGCATTGGACCAGAATAGAAGAAGAGGGGTGCCAAGGCGT 4549

Qy      1456 ACAATAAAAATGTAAAGAAGAGAAAGCATTCAAAGTGTGAGATCTCGGTGTGAGACTATTT 1515
             |||||||| | ||| ||  || || || |||||||| ||||| |||||||| ||||||||
Db      4550 ACACCAAGAGGGTGAAAGCAAAATTGTTTCAAGTTGGAGACTTGGTTTGGAAGACAATTT 4609

Qy      1516 TGCCTCTTCGGGGTAAAAGACAACAAGTTTAGTAAGTGGCCTCAAAATTGGGAGGGCCCA 1575
             ||||| |||||| ||   |||  ||||| |||||||| |||  |||  || ||||| ||
Db      4610 TGCCT-TTCGGGTACTCGATCGGAGCGTTCGGTAAGTCGTAAGTGGTCTCCTAGTTGGGAAGGTCCT 4668

Qy      1576 TGCAAGATTGTTAAAGTAATTGTTTTGGA------TTGACGGAGGCATTTCAAGGTGAT 1628
             | |||| ||||  |||||||||||| |||      ||||| ||| ||| |||| ||||
Db      4669 TATCGAATATGTGGCATTGTTCGAGGGACGACCGCATACTTTTTGGAGACACTTCAGGGGGAG 4728

Qy      1629 CATCTTACCTAGAGCTCTCAATGGAGGTGCTCGAAGACATATTACCCATGTGTATCGCAA 1688
             |||||||| | |||| || ||||||||||| |||||||  ||| ||| || ||| |||||
Db      4729 CGTTTTCAGCGAGCAATAACCGGGAAATAACTTGAGAAATACTTCCCGAGCGTTTGGCAA 4788

Qy      1689 GATGTTTA 1696
             || | |||
Db      4789 GACGCTTA 4796
```

```
RESULT 4
ADJ40495
ID   ADJ40495 standard; cDNA; 4797 BP.
XX
AC   ADJ40495;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Plant cDNA #1495.
XX
KW   Plant; gene; ss; transcription; plant genome augmentation; cereal;
KW   soybean; alfalfa; sunflower; canola; cotton; peanut; tobacco; sugar beet;
KW   maize; barley; sorghum; rice; wheat; crop plant; insecticide resistance;
KW   stress tolerance; salt tolerance; cold tolerance; drought tolerance;
KW   plant nutrition; apical dominance; dwarfism; early flowering; antiviral;
KW   antifungal.
XX
OS   Eukaryota.
XX
PN   US2004016025-A1.
XX
PD   22-JAN-2004.
XX
PF   26-SEP-2002; 2002US-00260238.
XX
PR   26-SEP-2001; 2001US-0325277P.
PR   26-SEP-2001; 2001US-0325448P.
PR   04-APR-2002; 2002US-0370620P.
XX
PA   (BUDW/)  BUDWORTH P.
PA   (MOUG/)  MOUGHAMER T.
PA   (BRIG/)  BRIGGS S P.
PA   (COOP/)  COOPER B.
PA   (GLAZ/)  GLAZEBROOK J.
PA   (GOFF/)  GOFF S A.
PA   (KATA/)  KATAGIRI F.
PA   (KREP/)  KREPS J.
PA   (PROV/)  PROVART N.
PA   (RICK/)  RICKE D.
PA   (ZHUT/)  ZHU T.
XX
PI   Budworth P, Moughamer T, Briggs SP, Cooper B, Glazebrook J;
PI   Goff SA, Katagiri F, Kreps J, Provart N, Ricke D, Zhu T;
XX
DR   WPI; 2004-190374/18.
XX
PT   New rice promoter, useful for manipulating crop plants to alter or
PT   improve phenotypic characteristics, e.g. produce large quantities of oil
PT   or proteins, resistance to insecticides, virus or fungi, stress tolerance
PT   or high nutritional value.
XX
PS   Claim 25; SEQ ID NO 1495; 230pp; English.

RESULT 5
ADJ41110/c
ID   ADJ41110 standard; cDNA; 2000 BP.
XX
AC   ADJ41110;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Plant cDNA #2110.
XX
```

```
Db   897  TTGGATGTTTGATTATCTAACTTTATGCAGATGTTGGCGTATAGAATCGGCATTTAAC  838
          ||| |||||  ||        ||    ||| |||||||||||||||||||||| |||||
Qy   643  GACCGCCAACAAGATCCCCACACTTAGCCTTCGTGCTCATCCTCGAGTAAGTTCAAGGA  702
          |||||||||| ||||||||||||| ||||||||||||||||||||||||||| ------
Db   837  AACCGCCAACAAAACCCCCACACTTAGCCTTCGTTTCCCTGAGTAAGGT------  784

Qy   703  CTAAGGTGGAACATCTCCTCAAATGTACGATGCCTGCATATAAGTATTCCAAGCCTCA  762
          |||  ||   |||||||||||||   ||||||||| ||| |||||||||||||||||||
Db   783  -TAATGTAGAATGAATTCTTCAAGAGATGAATCAAGCATATAAACCATTCACAACCATA  725

Qy   763  CCTATACATGTGAAC-TTTGAAGTGTCTACCACGCCATCTTGGGTGGTTGAGAAATGGAA  821
          ||||||| | || |||  |||||| ||||| |||||| ||||||||| ||||||||||||
Db   724  CCTGCACCTATGAACTTTTCAGTGGCTATCTTACCATCTTGAGTAATTGATAAATGGAA  665

Qy   822  CAGATCAGAATCCAGTCATCTTACCTCTCGTTGCCTCTTAGATAACTTGGGTTTTGTAAGGT  881
          ||||||||||||| || |||||||||||||   ||| ||| |||| |||||| ||| ||| |
Db   664  CGATCTTGATTCCAGTCCTTCTTGCTTCTCGTCAGACTTCATGGGATTTTTATGAACTTT  605

Qy   882  TTTCAA------ATTTAAAACATAGTCTTGCTCCTCCAAATGATTCTCTCATATAGCTCAA  935
          |||||||      | ||  |||   ||| | ||   || |||   |||||| |||| | | |
Db   604  TGCAAACTAGCTATGTTCCTCATATTTGACCTTCCTCTTGTCTCTCAAAGATTTATATT  545

Qy   936  TGTGTATGGTTTCTCACCAAGGCAATGTTTTGCCTCTTTTCATCCTACTTCTAATATTTC  995
          || |||| |||| |||||||    | || |||||||||||||||||||| ||||| |||||
Db   544  TTCTTTTGGATCCTCACCAAGACAAAGTGTGCCAGAATTTACCCCTA--TTCTATGGCTG  487

Qy   996  TTTGTGGAGCTTAGGGTAGGGAATGAAAAGGAAGCACTGGTGACATACCTACCATCATGTTACTA  1055
          ||||||||||||                        || |||| ||  ||||||| ||||  | |
Db   486  ATATGTGGAGCTT---GGTAGGGGATAAACTGGTGACATAAACTCCTACACATCATATGTATTGCAA  430

Qy  1056  AGTCAAAACCAAATCTGAGGAGAAGCA  1083
          |||||||  |||||||  ||    ||
Db   429  AGCCAAAAATATGGACTCAAGAAGATCA  402
```

```
RESULT 6
ADA71938
ID   ADA71938 standard; DNA; 2000 BP.
XX
AC   ADA71938;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Rice gene, SEQ ID 5263.
XX
KW   Plant; bacterial infection; fungal infection; viral infection; rice;
KW   gene; ds.
XX
OS   Oryza sativa.
XX
PN   WO200300898-A1.
XX
PD   03-JAN-2003.
XX
PF   22-JUN-2001; 2001WO-IB001105.
PR   22-JUN-2001; 2001WO-IB001105.
XX
PA   (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI   Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI   Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR   WPI; 2003-175290/17.
XX
PT   Identifying at least one gene involved in plant resistance or response to
PT   pathogenic infection for conferring resistance or tolerance to a plant to
PT   bacterial, fungal or viral infection by determining or detecting plant
PT   gene expression.
XX
PS   Claim 27; SEQ ID NO 5263; 899pp; English.
XX
```

```
KW   Plant; gene; ss; transcription; plant genome augmentation; cereal;
KW   soybean; alfalfa; sunflower; canola; cotton; peanut; tobacco; sugar beet;
KW   maize; barley; sorghum; rice; wheat; crop plant; insecticide resistance;
KW   stress tolerance; salt tolerance; cold tolerance; drought tolerance;
KW   plant nutrition; apical dominance; dwarfism; early flowering; antiviral;
KW   antifungal.
XX
OS   Eukaryota.
XX
PN   US200401025-A1.
XX
PD   22-JAN-2004.
XX
PF   26-SEP-2002; 2002US-00260238.
XX
PR   26-SEP-2001; 2001US-0325277P.
PR   26-SEP-2001; 2001US-0325448P.
PR   04-APR-2002; 2002US-0370620P.
XX
PA   (BUDW/) BUDWORTH P.
PA   (MOUG/) MOUGHAMER T.
PA   (BRIG/) BRIGGS S P.
PA   (COOP/) COOPER B.
PA   (GLAZ/) GLAZEBROOK J.
PA   (GOFF/) GOFF S A.
PA   (KATA/) KATAGIRI F.
PA   (KREP/) KREPS J.
PA   (PROV/) PROVART N.
PA   (RICK/) RICKE D.
PA   (ZHUT/) ZHU T.
XX
PI   Budworth P, Moughamer T, Briggs SP, Cooper B, Glazebrook J;
PI   Goff SA, Katagiri F, Kreps J, Provart N, Ricke D, Zhu T;
XX
DR   WPI; 2004-190374/18.
XX
PT   New rice promoter, useful for manipulating crop plants to alter or
PT   improve phenotypic characteristics, e.g. produce large quantities of oil
PT   or proteins, resistance to insecticides, virus or fungi, stress tolerance
PT   or high nutritional value.
XX
PS   Claim 60; SEQ ID NO 2110; 230pp; English.
XX
CC   The invention relates to plant nucleotide sequences that direct seed-,
CC   leaf- and/or stem-, panicle-, root- or pollen-specific or -preferential
CC   or constitutive transcription of an operatively linked nucleic acid
CC   segment. The invention also relates to a method for augmenting a plant
CC   genome and a method of identifying a gene, where its expression is
CC   altered in the seed, leaf, stem, panicle, pollen, root or is constitutive
CC   in a plant cell. The plant is a cereal, e.g. soybean, alfalfa, sunflower,
CC   canola, cotton, peanut, tobacco or sugar beet, preferably maize, barley,
CC   sorghum, rice or wheat. The polynucleotides and the polypeptides they
CC   encode are useful for manipulating crop plants to alter or improve
CC   phenotypic characteristics, to produce large quantities of oil or
CC   proteins, to incur resistance to insecticides, viruses or fungi, and to
CC   incur stress tolerance (e.g. salt, cold or drought) to ensure the plants
CC   have a high nutritional value with reduced apical dominance or dwarfism,
CC   early flowering or altered metabolic pathways. This sequence represents a
CC   plant nucleic acid of the invention. Note: The sequence data for this
CC   patent did not form part of the printed specification but was obtained in
CC   electronic format directly from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 2000 BP; 629 A; 374 C; 421 G; 576 T; 0 U; 0 Other;
```

```
Query Match             4.3%;  Score 114.4; DB 12; Length 2000;
Best Local Similarity   57.6%;  Pred. No. 4.5e-20;
Matches  327;  Conservative   0;  Mismatches  221;  Indels  20;  Gaps   6;

Qy   524  CTACAAATGATAAATCACCAAAACTCATGGAGCTTGCTAGTTATAAACTCTAATTCTAAG  583
          ||||||||||||||||    |||||| |||| |||| |||||||||| ||| |||||
Db   957  CTGCATACAAATATATACCAACACTTGTGGAATATATGGAATATGAGATTAAGCACTTATCACTTATT  898

Qy   584  TTTGGTGTTTTATTGAGTGGATTTTGTGTGAAAGTTGGTGGT-TAGAAATAGGAGTTAAG  642
```

```
CC   The present invention relates to a method (M1) for identifying genes
CC   involved in plant resistance or response to pathogenic infection. M1
CC   comprises identifying a gene whose expression is significantly altered in
CC   the incompatible interaction of plant gene expression relative to
CC   expression of the gene in an uninfected plant, in a mutant plant that
CC   does not express a gene associated with response to pathogenic infection,
CC   or in a corresponding incompatible or compatible interaction. (M1) is
CC   useful for conferring resistance to resistance or tolerance to a plant to
CC   bacterial, fungal or viral infection. The present sequence was used to
CC   illustrate the invention.
XX
SQ   Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match              2.1%;    Score 55.4;  DB 8;  Length 2000;
Best Local Similarity   9.9%;     Pred. No. 0.00064;
Matches   75;  Conservative 351;  Mismatches 327;  Indels    4;  Gaps   3;

Qy          45 CGGGCTGGTCTGCGACAGCTAGAGGCGGCCACCGCGTCCTAGCTTCCTCCAACTTCTCGTC 104
               ::: : :::::: : |:::::: |::::: |||||:|| :::: |||: : :|::
Db         272 SRSYGWYSWSYKWWMCTAYKKSYYSRWCYMYRGGWRGATRYWRGYMSRMAMMYKKMYW 331

Qy         105 GGAGATCCCTTCAGGGATGCCAATGCCACCGCCCTAGTCAACCTGCGGGAGCTGGAG 164
               ::: : ||:|: |||:|| : ||:: :: ::: ::::::: || : :::::
Db         332 YRGYKGMKRGWWAGRMWMRSMCRWSKACYMRWRWRMRWMTRRRRWAKKSSRTSRRKKRKWC 391

Qy         165 CTTCGCCAGGGTCAGAGCTGCGGCAGCACCCTGGTAGACCGCATTCCTGATGACCCGCGG 224
               :::: :::||: ::::: ::: :||:: : :::|| :::::|||::| : :::::
Db         392 MRKRKYKRMRGYSMRSCKRARWMRKCRSRGRAWKMGCRGMTCRMKSYGMMRWKSWKRMA 451

Qy         225 GGTGCGCTCCATGAAGAAGTTGCATTCGCCCAACCAAGTCGAGTGGGTCGCCTGGAGGGGG 284
               :::: :|:: : : : :|||::: || ::: :: : || :::: ::::
Db         452 SKYKWMSRMYRWRKKKCSRTTMWGKTRGGMMGTMGRCRYK-KRSGMKRKCRRRRWGRMYR 510

Qy         285 CGGGGAAGCAAAACGTTGCATGCACCTAGCGCCCTGGCAGCGAGCTCCTGTAGTATCACC 344
               :: : : ::: ::: : :: :::
Db         511 MRWKRYYMSARYTMRYCARKKYSYSAARKARCWYRGKGYYWAGMWMKRYKRMYMYKMWW 570

Qy         345 TGGCTGCCTCCAGCTCATGCTCGCAAGCCTCCAGGGCGGCCCCGCAGTGCTCCAACACT 404
               :::: :::: : :|||: :: :: ::: ::|||::: : : :::|::
Db         571 YKRKYSKCSWYCKWSYYASCMKSARKAGAKMCKRSKSMAWSKSMRSSRRKCRKCASKRSSA 630

Qy         405 TTCGCCCTCCTCCTACAGCTCCCTTCCACATGCAGTCGTGCTCCGCACGCACCTTCTCCACC 464
               :::: :::||| ::: | : :: : ::::::::: : : :::: ::::
Db         631 KRYAMMGGMTSGSRMSRWKSYTCYWRKWGSMKSTCTWMYMSKYTYAKYGSYWRYRYRAW 690

Qy         465 TTTTTTACTCTTTTTCTTTTCTTTCTTGGCCCATCTTTGTATTTTCACAAAATGTCCCCC 524
               ::: :::: ::|::|:|| :: ::: ::||::: :: : ::: ::
Db         691 CMYMWRWYYYRYRSYMTYMAWYISST--RMAMTGMKYSGRYWTSWYKYCKCSWKYRSMWY 748

Qy         525 TACAAATGATAAATCACCAAAACTCATGGAGCTTGCTAGTTATAAACTCTAATTCTAAGT 584
               :: :::: ::: ::: : :||: | :|: :::: :|: :::
Db         749 YWSWWWAKTWKMWRRYATRMWWMWYRYSMKWYTWCTMWGYWVWYWVRTYWMKMRYMWYKCTK 808

Qy         585 TTGGTGTTTATTTGAGTGGATTTTGTGTGAAAGTTGGTGTTAGAAAATAGGAGGTTAAGGA 644
               |||:| :::|: :::||: |:||::|:: : :::||::|: ::|
Db         809 TYWYWSATYWTGTWAAWMAKTKMRMGMTGAKTRGRARKARYWMKWATWCATKRWMTKGK 868

Qy         645 CCGCAACAAGATCCCCCACTTAGCCCTTTGCTCATCCTCGAGTAAAGTTCAAGGACT 704
               : :: ::: ::::: :: : : : : :::|| : ::|: |
Db         869 GAKWAWTWMAKAW-RKKYYWSWMRAWYYYYKTRRTRYKTCWWKARWGSWAYWRMWWKGSAK 927

Qy         705 AAGGTGGAACATCTCCTCAAATGTACAGATGCCTGCATATAAGTTATTCCAAGCCTCACC 764
               :::: : : :::::|||::::|::: : : : ::
Db         928 MWMWNKGGGRWGWTKYWYWVYCTTWKMACGRATKKYMCCAGWWAMYSYSWTRTYWMRTWRWM 987

Qy         765 TATACATGTGAACTTGAAGTGTCTACCACGCCATCT 801
               ||: : ::: :::::||:::||:|| :| ||
Db         988 ASSRTAKRMARMWMKTRAWSKSYARAYWKMAGCACCT 1024

RESULT 7
AAX83007
ID   AAX83007 standard; DNA; 51259 BP.
```

```
XX
AC   AAX83007;
XX
DT   31-AUG-1999  (first entry)
XX
DE   Partial mouse WRN genomic sequence #3.
XX
KW   Mouse; WRN; Werner's syndrome; detection; diagnosis; autosomal;
KW   recessive disorder; phenotype; ss.
XX
OS   Mus musculus.
XX
PN   WO9724435-A1.
XX
PD   10-JUL-1997.
XX
PF   30-DEC-1996;          96WO-US020785.
XX
PR   29-DEC-1995;          95US-0009409P.
PR   29-DEC-1995;          95US-00580539.
PR   30-JAN-1996;          96US-00108035P.
PR   30-JAN-1996;          96US-00594242.
PR   12-APR-1996;          96US-00612175.
XX
PA   (DARW-) DARWIN MOLECULAR CORP.
XX
PI   Oshima J, Fu Y, Yu C, Mulligan J, Schellenberg GD;
XX
DR   WPI; 1997-363671/33.
XX
PT   Isolated nucleic acid molecule encoding the WRN gene product - useful for
PT   detection and treatment of Werner's syndrome, and related diseases.
XX
PS   Claim 1; Fig 7; 153pp; English.
XX
CC   This sequence represents a fragment of the genomic sequence containing
CC   the coding region for the mouse WRN gene (AAX83004). The corresponding
CC   human gene (AAX83001) encodes a protein related to Werner's syndrome. The
CC   products can be used for the detection and treatment of Werner's syndrome
CC   (WS), an autosomal recessive disorder with a complex phenotype, as well
CC   as related diseases
XX
SQ   Sequence 51259 BP; 14533 A; 9635 C; 10266 G; 16825 T; 0 U; 0 Other;

Query Match              2.0%;    Score 53.4;  DB 2;  Length 51259;
Best Local Similarity  54.3%;     Pred. No. 0.013;
Matches  108;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps   0;

Qy         314 CGGCCTGGCAGCGAGCTCCTGTAGTATCACCTGGCTGCCGCCTCCAGCTCATGCTCGCAAGC 373
               ||||||||||| |||||| ||||||||||||||||| ||||| |||||||||||||||||||
Db         107 CTCCTTCTCCTTCTCCTTCTCCTCGTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTC 166

Qy         374 CTCCAGGGCGGCCCCGCAGTGCTCCAACACTTTCGCCCTCCTCCTCCTCCTCCTCCTCCACAT 433
               ||||||||| |||| |||| |||||| |||| ||||||||||||||||||||||||||||||
Db         167 CTGCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCC 226

Qy         434 GCAGTCGTGCTCCGCACGCACCTTCTCCACCTTTTACCTTTTTTCTTTTCTTTTCTTGG 493
               |||| ||||||||||||||||||||| |||||||||||| ||||||||||||||||||||
Db         227 CTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCATGTATTTGTGTGTTTTAG 286

Qy         494 CCCATCTTTGTATTTTCA 512
               |||||||||||| ||||
Db         287 ACATTCTGTGTTTTACTCA 305

RESULT 8
ADQ97067
ID   ADQ97067 standard; DNA; 72352 BP.
XX
AC   ADQ97067;
XX
DT   07-OCT-2004  (first entry)
XX
```

```
DE   Mouse cancer associated sequence MD11-040, SEQ ID 43.
XX
KW   Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Mouse; ds.
XX
OS   Mus musculus.
XX
PN   WO200406030304-A2.
XX
PD   22-JUL-2004.
XX
PF   22-DEC-2003; 2003WO-US041389.
XX
PR   27-DEC-2002; 2002US-00310773.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW, Malandro MS;
XX
DR   WPI; 2004-543781/52.
XX
PT   New isolated cancer associated nucleic acids comprising at least 10
PT   contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT   cancers such as leukemia and lymphoma.
XX
PS   Claim 1; SEQ ID NO 43; 199pp; English.
XX
CC   The present invention relates to cancer associated sequences (ADQ97025-
CC   ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC   treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic formate directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 72352 BP; 15474 C; 15337 G; 19982 T; 0 U; 1901 Other;

Query Match             1.9%;  Score 52.2;  DB 12;  Length 72352;
Best Local Similarity  56.8%;  Pred. No. 0.034;
Matches  96;  Conservative   0;  Mismatches  73;  Indels   0;  Gaps   0;

Qy     328 GCTCCTGTAGTATCACCTGGCTGCGTCCGCTCCAGCTCATGCTCGCAAGCCTCCAGGGCGGCCC 387
           ||||||  |||||||  || |||  || ||||| ||||||  ||||
Db   63505 GCTCCTGCTCCTGCTCCTGCTCCTGCTCCTGCTCCTGCTCCTGCTCCTGCTCCTCCTCCTCCT 63564

Qy     388 GGCAGTGCTCCAACACTTTCGCCTCCTCCTACAGCTCCTTCCACATGCAGTCGTGCTCCG 447
           |||| ||||||||     | ||| || || ||   ||    ||||
Db   63565 GCTCCTGCTCCTGCTCCTGCACCTGCTCCTCCTCCTGCTCCTCCTGCTCCTGCTCCTGCT 63624

Qy     448 CACCGACCTTCTCCACCTTTTACCTCTTTTTCTTTTCTTTTCTTGGCCC 496
           ||||||||  | ||||| |     |||   || ||  ||||
Db   63625 CCTGCTCCTGCTCCCCTCTCTCTGCTCCCCTCTCCTCTCTCAGCTC 63673

RESULT 9
AAV72244
ID   AAV72244 standard; DNA; 3718 BP.
XX
AC   AAV72244;
XX
DT   24-MAY-1999 (first entry)
XX
DE   G. max SBP2 DNA 5' regulatory region.
XX
KW   SBP1; sucrose binding protein; SBP2; sucrose uptake; transgenic plant;
KW   seed; carbohydrate content; soybean; ds.
XX
OS   Glycine max.
XX
PN   WO9853086-A1.
XX
PD   26-NOV-1998.
XX
PF   21-MAY-1998; 98WO-US010465.
XX
```

```
PR   22-MAY-1997;  97US-0047568P.
XX
PA   (UNIW ) UNIV WASHINGTON STATE RES FOUND.
XX
PI   Grimes HD, Chao WS;
XX
DR   WPI; 1999-070155/06.
XX
PT   New modified plant sucrose binding proteins - used to develop transgenic
PT   plants which can have enhanced or decreased sucrose uptake activity in
PT   developing seeds.
XX
PS   Claim 17; Page 42-43; 58pp; English.
XX
CC   This sequence represents the Glycine max sucrose binding protein SBP2 DNA
CC   5' regulatory region. The SBP2 protein is used in a method resulting in
CC   the production of a modified plant sucrose binding protein (SBP) which
CC   has a modified amino acid sequence compared to a corresponding wild-type
CC   SBP, and where expression of the modified SBP in a yeast assay system
CC   confers enhanced sucrose uptake compared to the corresponding wild-type
CC   SBP. The products of the invention can be used for producing transgenic
CC   plants which have modified sucrose uptake activity, particularly in
CC   developing seeds. Enhanced sucrose uptake activity in developing seeds
CC   may be desirable where it is an advantage to increase the carbohydrate
CC   content of the seed (e.g. where the seed is the primary plant material
CC   harvested, such as soybean). In contrast, decreased sucrose uptake
CC   activity in seeds might be desirable where the vegetative material of the
CC   plant is harvested. The SBP regulatory regions confer specific or
CC   enhanced expression in developing seeds and so may be used to express any
CC   transgene in developing seeds
XX
SQ   Sequence 3718 BP; 1145 A; 656 C; 644 G; 1273 T; 0 U; 0 Other;

Query Match             1.9%;  Score 51.8;  DB 2;  Length 3718;
Best Local Similarity  96.4%;  Pred. No. 0.0086;
Matches  53;  Conservative   0;  Mismatches   2;  Indels   0;  Gaps   0;

Qy     7 TAATACGACTCACTATAGGGACACGCGTGGTCGACGGCCCGGCGTGGTCTGCGACA 61
         |||||||||||||||||||||| ||||||||||||||||||||| ||||
Db    99 TAATACGACTCACTATAGGGACACGCGTGGTCGACGGCCCGGCGTGGTCTGAGAAA 153

RESULT 10
AB257867
ID   AB257867 standard; DNA; 1150 BP.
XX
AC   AB257867;
XX
DT   14-APR-2003 (first entry)
XX
DE   Porcine endogenous retrovirus flanking sequence.
XX
KW   PERV; knockout animal; pig; xerotransplantation; organ transplant; ss.
XX
OS   Porcine endogenous retrovirus.
XX
FH   Key            Location/Qualifiers
FT   misc_feature   1..73
FT                  /tag= a
FT                  /note= "vector"
FT   primer_bind    53..73
FT                  /tag= b
FT   primer_bind    617..636
FT                  /tag= c
FT   LTR            872..900
FT                  /tag= d
FT   primer_bind    872..900
FT                  /tag= e
FT   primer_bind    1051..1071
FT                  /tag= f
FT   misc_feature   1072..1150
FT                  /tag= f
FT                  /note= "vector"
```

XX
PN   WO2003002746-A2.
XX
PD   09-JAN-2003.
XX
PF   28-JUN-2002; 2002WO-EP007159.
XX
PR   29-JUN-2001; 2001US-0302133P.
XX
PA   (NOVS) NOVARTIS AG.
PA   (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Herring CT, Langford G, Quinn G, Scobie L;
XX
DR   WPI; 2003-210279/20.
XX
PT   An isolated polypeptide useful for screening pigs for the presence of a
PT   porcine endogenous retrovirus (PERV) and providing a pig modified not to
PT   express a PERV.
XX
PS   Claim 1; Page 57-58; 63pp; English.
XX
CC   The present sequence is that of a claimed flanking sequence of porcine
CC   endogenous retrovirus (PERV). Flanking sequences are DNA sequences
CC   surrounding a PERV which represent a unique molecular signature that can
CC   be used to characterise PERV integration sites and identify PERVs. The
CC   present sequence is one of eight claimed PERV polynucleotides (see
CC   AB257861-68) useful for screening a pig for the presence of PERV. The
CC   present sequence is useful for screening a pig for the presence of PERV. The
CC   screening provides for the elimination of donors with known proviruses. A
CC   pig modified not to express a selected PERV is obtained by identifying a
CC   PERV using the claimed polynucleotides, and knock-out or inactivation of
CC   the PERV by homologous recombination. The organs, tissues and cells of
CC   the modified pig are suitable for use in xerotransplantation
XX
SQ   Sequence 1150 BP; 333 A; 265 C; 247 G; 305 T; 0 U; 0 Other;

Query Match   1.9%; Score 51; DB 8; Length 1150;
Best Local Similarity 91.5%; Pred. No. 0.0075;
Matches 54; Conservative   0; Mismatches   5; Indels   0; Gaps   0;

Qy   7 TAATACGACTCACTATAGGGCACGCGTGGTCGACGGCCCGGGCGTCTGCGACAGCTA 65
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54 TAATACGACTCACTATAGGGCACGCGTGGTCGACGGCCCGGGCGTCTGCGACAGCTA 112

RESULT 11
AB257861
ID   AB257861 standard; DNA; 2860 BP.
XX
AC   AB257861;
XX
DT   14-APR-2003 (first entry)
XX
DE   Porcine endogenous retrovirus flanking sequence.
XX
KW   PERV; knockout animal; pig; xerotransplantation; organ transplant; ss.
XX
OS   Porcine endogenous retrovirus.
XX
FH   Key   Location/Qualifiers
FT   misc_feature   1..21
FT   /*tag= a
FT   /note= "vector"
FT   primer_bind   2536..2556
FT   /*tag= b
FT   LTR   2661..2860
FT   /*tag= d
FT   primer_bind   2661..2678
FT   /*tag= c
FT   primer_bind   2839..2858
FT   /*tag= e
XX
PN   WO2003002746-A2.

XX
PD   09-JAN-2003.
XX
PF   28-JUN-2002; 2002WO-EP007159.
XX
PR   29-JUN-2001; 2001US-0302133P.
XX
PA   (NOVS) NOVARTIS AG.
PA   (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Herring CT, Langford G, Quinn G, Scobie L;
XX
DR   WPI; 2003-210279/20.
XX
PT   An isolated polypeptide useful for screening pigs for the presence of a
PT   porcine endogenous retrovirus (PERV) and providing a pig modified not to
PT   express a PERV.
XX
PS   Claim 1; Page 45-47; 63pp; English.
XX
CC   The present sequence is that of a claimed flanking sequence of porcine
CC   endogenous retrovirus (PERV). Flanking sequences are DNA sequences
CC   surrounding a PERV which represent a unique molecular signature that can
CC   be used to characterise PERV integration sites and identify PERVs. The
CC   present sequence is one of eight claimed PERV polynucleotides (see
CC   AB257861-68) useful for screening a pig for the presence of PERV. The
CC   screening provides for the elimination of donors with known proviruses. A
CC   pig modified not to express a selected PERV is obtained by identifying a
CC   PERV using the claimed polynucleotides, and knock-out or inactivation of
CC   the PERV by homologous recombination. The organs, tissues and cells of
CC   the modified pig are suitable for use in xerotransplantation
XX
SQ   Sequence 2860 BP; 691 A; 492 C; 533 G; 1144 T; 0 U; 0 Other;

Query Match   1.9%; Score 51; DB 8; Length 2860;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 51; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   7 TAATACGACTCACTATAGGGCACGCGTGGTCGACGGCCCGGGCGTCTGC 57
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db   2 TAATACGACTCACTATAGGGCACGCGTGGTCGACGGCCCGGGCGTCTGC 52

RESULT 12
AB257865/c
ID   AB257865 standard; DNA; 1976 BP.
XX
AC   AB257865;
XX
DT   14-APR-2003 (first entry)
XX
DE   Porcine endogenous retrovirus flanking sequence.
XX
KW   PERV; knockout animal; pig; xerotransplantation; organ transplant; ss.
XX
OS   Porcine endogenous retrovirus.
XX
FH   Key   Location/Qualifiers
FT   misc_feature   1..54
FT   /*tag= a
FT   /note= "vector"
FT   primer_bind   55..75
FT   /*tag= b
FT   LTR   76..504
FT   /*tag= c
FT   primer_bind   486..504
FT   /*tag= d
FT   primer_bind   1676..1695
FT   /*tag= e
FT   misc_feature   1704..1976
FT   /*tag= f
FT   /note= "vector"
FT   primer_bind   1704..1733

FT   /*tag= f
XX
PN   WO200300746-A2.
XX
PD   09-JAN-2003.
XX
PF   28-JUN-2002; 2002WO-EP007159.
XX
PR   29-JUN-2001; 2001US-0302133P.
XX
PA   (NOVS) NOVARTIS AG.
PA   (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Herring CT, Langford G, Quinn G, Scobie L;
XX
DR   WPI; 2003-210279/20.
XX
PT   An isolated polypeptide useful for screening pigs for the presence of a
PT   porcine endogenous retrovirus (PERV) and providing a pig modified not to
PT   express a PERV.
XX
PS   Claim 1; Page 54-55; 63pp; English.
XX
CC   The present sequence is that of a claimed flanking sequence of porcine
CC   endogenous retrovirus (PERV). Flanking sequences are DNA sequences
CC   surrounding a PERV which represent a unique molecular signature that can
CC   be used to characterise PERV integration sites and identify PERVs. The
CC   present sequence is one of eight claimed PERV polynucleotides (see
CC   AB257861-68) useful for screening a pig for the presence of PERV. The
CC   screening provides for the elimination of donors with known proviruses. A
CC   pig modified not to express a selected PERV is obtained by identifying a
CC   PERV using the claimed polynucleotides, and knock-out or inactivation of
CC   the PERV by homologous recombination. The organs, tissues and cells of
CC   the modified pig are suitable for use in xerotransplantation
XX
SQ   Sequence 1976 BP; 612 A; 425 C; 406 G; 533 T; 0 U; 0 Other;

Query Match   1.9%; Score 50.8; DB 8; Length 1976;
Best Local Similarity   96.3%; Pred. No. 0.011;
Matches   52; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   7 TAATACGACTCACTATAGGGCACGCGTGGTCGACGGCCCGGGCTGGTCTGCGAC 60
          |||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   1900 TAATACGACTCACTATAGGGCACGCGTGGTCGACGGCCCGGGCTGGTCTGTGGC 1847

RESULT 13
AAK45604
ID   AAK45604 standard; DNA; 267 BP.
XX
AC   AAK45604;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human bone marrow expressed single exon probe SEQ ID NO: 20161.
XX
KW   Human; bone marrow expressed exon; gene expression analysis; probe;
KW   microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157276-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000668.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-0002Y263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-488900/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human bone marrow.
XX
PS   Example 4; SEQ ID NO 20161; 658pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   bone marrow. They can be used to measure gene expression in bone marrow
CC   samples, which may enable the improved diagnosis and treatment of cancers
CC   such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC   the probes of the invention
XX
SQ   Sequence 267 BP; 3 A; 151 C; 4 G; 109 T; 0 U; 0 Other;

Query Match   1.9%; Score 50; DB 4; Length 267;
Best Local Similarity   56.8%; Pred. No. 0.0063;
Matches   92; Conservative   0; Mismatches   70; Indels   0; Gaps   0;

Qy   341 CACCTGCGTCGCCTCCCAGCTCATGCTGCAAGCCTCGACAGCCGGCCGGCCAGTGCTCCAA 400
                    |||||||| |||| | ||| |||| |||||||||||||||||||
Db   73 CTCCTCCTTCTCCTCCTCCTACAGCTCCTCCCTC 132

Qy   401 CACTTTCGCCTCCTCCTACAGCTCCTTCCACATGCAGTCAGTGCTGTCCGCACGCACCTTCTC 460
          || ||||||||||| |||| |||| |||| ||||||||
Db   133 CTCCTTCTCCTCCTCCTC 192

Qy   461 CACCTTTTACTCTTTTCTTTCTCTTTTTCTTGGCCCATCTTT 502
          | ||||||||||||| ||||||||||||||
Db   193 CTCCTCCTCCTC CCACCTGT 234

RESULT 14
AAK19599
ID   AAK19599 standard; DNA; 267 BP.
XX
AC   AAK19599;
XX
DT   05-NOV-2001 (first entry)
XX
DE   Human brain expressed single exon probe SEQ ID NO: 19590.
XX
KW   Human; brain expressed exon; gene expression analysis; probe; microarray;
KW   Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;
KW   ss.
XX
OS   Homo sapiens.
XX
PN   WO200157275-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000667.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0236359P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX

DR   WPI; 2001-483446/52.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   brains.
XX
PS   Example 4; SEQ ID NO 19590; 650pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   brain. They can be used to measure gene expression in brain cell samples,
CC   which may enable the diagnosis and improved treatment of nervous system
CC   diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC   epilepsy and cancers. The present sequence is one of the probes of the
CC   invention
XX
SQ   Sequence 267 BP; 3 A; 151 C; 4 G; 109 T; 0 U; 0 Other;

Query Match   1.9%;   Score 50; DB 4; Length 267;
Best Local Similarity   56.8%;   Pred. No. 0.0063;
Matches   92;   Conservative   0;   Mismatches   70;   Indels   0;   Gaps   0;

Qy   341 CACCTGGTGCGCCTCCAGCTCATGCTCGGCAAGCCTCCAGGGCGGCCCGGCAGTGCTCCAA 400
Db   73 CTCCTCCTTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTTCTCCTCCTC 132

Qy   401 CACTTTCGCCTCCTCCTACAGCTCCTTCCACATGCAGTCGTTCCGCGACGACCTTCTC 460
Db   133 CTCCTTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTTCTCCTCCTTCTC 192

Qy   461 CACCTTTTACTCTTTTCTTTCTTTCTTTCTTGGCCCATCTT 502
Db   193 CTCCTCCTCCTCCTCCTCCTCCTCCCCTCCTCCTCCACCTGT 234

RESULT 15
ABS45294
ID   ABS45294 standard; DNA; 267 BP.
XX
AC   ABS45294;
XX
DT   25-FEB-2003   (first entry)
XX
DE   Human liver single exon probe, SEQ ID No 20284.
XX
KW   Human; single exon nucleic acid probe; liver; cirrhosis;
KW   hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
KW   coronary heart disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157273-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000664.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-488898/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human adult liver.
XX

PS   Claim 4; SEQ ID NO 20284; 658pp; English.
XX
CC   The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC   measuring human gene expression in a sample derived from human adult
CC   liver, comprising one of 13109 defined nucleotide sequences given in the
CC   specification (or complements/ fragments). The probe hybridises at high
CC   stringency to a nucleic acid molecule expressed in the human adult liver.
CC   (I) may be used for predicting, measuring and displaying gene expression
CC   in samples derived from human adult liver. The genes identified may be
CC   involved in genetic liver diseases such as cirrhosis,
CC   hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC   associated with coronary heart disease. ABS25011-ABS51005 represent human
CC   liver single exon nucleic acid probes of the invention. Note: The
CC   sequence information for this patent does not appear in the printed
CC   specification but was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 267 BP; 3 A; 151 C; 4 G; 109 T; 0 U; 0 Other;

Query Match   1.9%;   Score 50; DB 4; Length 267;
Best Local Similarity   56.8%;   Pred. No. 0.0063;
Matches   92;   Conservative   0;   Mismatches   70;   Indels   0;   Gaps   0;

Qy   341 CACCTGGTGCGCCTCCAGCTCATGCTCGGCAAGCCTCCAGGGCGGCCCGGCAGTGCTCCAA 400
Db   73 CTCCTCCTTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTTCTCCTCCTC 132

Qy   401 CACTTTCGCCTCCTCCTACAGCTCCTTCCACATGCAGTCGTTCCGCGACGACCTTCTC 460
Db   133 CTCCTTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTTCTCCTCCTTCTC 192

Qy   461 CACCTTTTACTCTTTTCTTTCTTTCTTTCTTGGCCCATCTT 502
Db   193 CTCCTCCTCCTCCTCCTCCTCCTCCCCTCCTCCTCCACCTGT 234

Search completed: May 9, 2005, 02:15:24
Job time :1420 secs

THIS PAGE LEFT BLANK

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:       May  8, 2005, 21:32:08 ; Search time 451 Seconds
                                (without alignments)
                              9759.611 Million cell updates/sec

Title:          US-10-751-550-1
Perfect score:  2690
Sequence:       2690  1 tctagataatacgactcact........ccataactagtacaccatgg 2690

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters:      2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :  Issued_Patents_NA:*
            1: /cgm2_6/ptodata/1/ina/5A_COMB.seq:*
            2: /cgm2_6/ptodata/1/ina/5B_COMB.seq:*
            3: /cgm2_6/ptodata/1/ina/6A_COMB.seq:*
            4: /cgm2_6/ptodata/1/ina/6B_COMB.seq:*
            5: /cgm2_6/ptodata/1/ina/PCTUS_COMB.seq:*
            6: /cgm2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                            SUMMARIES
```

(table content)

```
Best Local Similarity   8.1%;  Pred. No. 5.7e-05;
Matches  29; Conservative 186; Mismatches 141; Indels   0; Gaps   0;

Qy    300 TTGCATGCACCTAGCGCCCTGCGCAGCGAGCTCCTGTAGTATCACCTGGTCGGTCGCCTCCAGC  359
                ::: :::::          :::::::  YYYYYYYYYYYYYYYYYYYYYYYYYYY
Db   1143 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY  1202

Qy    360 TCATGCTCGCAAGCCTCCAGGCGGCGGCCCGGCAGTGCTCCAACACTTTCGCCTCCTCCTAC  419
          YYYYYYY :::::  ::::::: YYYYYYYYYYYYYYYYYYYYYYYYYY :::::::::::
Db   1203 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY  1262

Qy    420 AGCTCCTTCCACATGCAGTCGTGCTCCGCACGCACCTTCTCCACCTTTTACTCTTTTCT  479
          YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY
Db   1263 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY  1322

Qy    480 TTCTCTTTTCTTGGCCCATCTTGGTATTTTCACAAATGTCCCCCTACACAAATGATAAATC  539
           :::::::::::::  :::::  YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY
Db   1323 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY  1382

Qy    540 ACCAAAACTCATGGAGCTTGCTAGTTATAAACTCTAATCCTAAGTTGGTGTTTATTTGA  599
          YYYYYYYYYYYYYYYYYYYYYYYYY :::  ::: ::::  YYYYYYYYYYYYYY
Db   1383 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCAAA  1442

Qy    600 GTGGATTTTGTTGTGAAAGTTGGTGGTTGAGAAATAGGAGTTAAGGACCGCCAACAAG  655
          |    |          :: :::          ||| ||:
Db   1443 TCTTCTATCTCTTTAAACTACTTGCATAGATAGGTAATTACAGTGATGCGTACATG  1498
```

```
Query Match  2.0%; Score 54; DB 4; Length 1141;
Best Local Similarity  10.1%;  Pred. No. 6.7e-05;
Matches 101; Conservative 352; Mismatches 543; Indels   0; Gaps   0;

Qy   1305 TTTACCTGTTGAGGTAAATCTGAACGCTAATAAATCGGCTAAGCAAAATAATTTATCACC  1364
           :  :|::::  :::  ::: ::::     :  :: :::          ||::
Db   1071 TNRRWCMRTYAMRTWYTRSNANWSCATKBMWTWTMKWYATKYRTAWYAMWCAWRNNNMWCA  1012

Qy   1365 TGTTGATTCTCAACAATTTGATGATGGACAAATATTGATGAGGTGACTGACAAATCATTGAA  1424
          :  ::  ::: :       ::  : :::          ::  ||| |:::
Db   1011 TNGYAKSCATNNAMWYATTRWAAYAAAKWARWAGNNRMRYGAAAGNKWGCMAAMATMGBW  952

Qy   1425 GGCTTTAAAGGAGGAGAAGGAGAAGGATAAAATCTACAATAAAAATGTAAAGGAAGAAGCATTC  1484
          :|:     ::::  ::: ::|: :::::          |::::::
Db    951 WADTFAGKMCNNNNNWTTDVRRMAMKAKNNNNNAYWTACYNRAATNNKMATHWMKWTHG  892

Qy   1485 AAAGTGTAGATCTGGTGTGAAGACTATTTTGCCTCTTGGGGTAAAAGACACAAGTT  1544
          |:: :|::||:   ::|::::   :: ::|::
Db    891 AHSKRRTRHHTRTCRRTKYNNNNNNARTVYWYHHAARRWMNAWWTRTNNNNNNNNNACR  832

Qy   1545 TAGTAAGTGGCCTCAAAATTGGAGGGGCCCATGCAAGATTGTTAAAGTAATTGTTTTGGA  1604
```

```
Db    831 NTRTWABWKHSWCNNNNNNNNNNNNNNTWCHYTTANABBCYRANNNNAAARMARTCNNY  772

Qy   1605 TTGACGGAGGCATTCAAGGTGATCATCACCTAGAGCTCTACCTAGAGCTCCAATGGGAGGTGCTCGAAG  1664
           :|:: |:::                   ::::  ::
Db    771 MHAAVTTHTDWCYKTWMNTWYWDMWTTMBTTTTRNMTTSTNMTNNNNNNNWACTNNNNN  712

Qy   1665 ACATATTACCCATGTGTATGGCAAAGATGTTTAGCTAGTAACTGACTGATAGTGTAAACGA  1724
          :: ::|::|:     ||:  :::  ::
Db    711 NMWKAYYAHATNNWGCWWNNTDARRITNNTTVMRRRWMTNTKTRWYSTTRRHHYTGATNNN  652

Qy   1725 TCTCCAATGGGCAAGACACATATTACCTAAGGCCAGGCTGGTTTTGCAAGTTCGAGTAGG  1784
          ::: ::::    :: :|::   ::::::  :: |||
Db    651 NNNNNNNNNNNSCCTCTRNMTMRWTMKGDGMTVRKKVKWRDTTCTYVDWADSWVWWY  592

Qy   1785 ATATAGAGATTCTCGTGCGAGTTGTAAACGATTCTCCAATGGGGCAAGACATCCTACACCCTA  1844
          |      ::::|: ::     :::::  ::::
Db    591 ANWMRCRDVTYTRNNTYCKSYAHSVYWYWSNNAMWYRRYSARNWSSMARWTTRNNWMSGB  532

Qy   1845 TATATAGTGAAGGGGCAGTAGCTGATTGAGAATCAATCAAGCACAATAATATTAT  1904
           :: |||| |::|     :|:||:     ::|  :::
Db    531 VRMRWAGTMWWRHWNNNNTDTRYWWKRWARBTTTVYDSMCNAKSMRGNNWRAMKMWW  472

Qy   1905 TAATTTTTATTCAAACCCAATTTTTTCCTTTTCCAACCCTAATTATAGTTTTCCTTTT  1964
          ||||           ::::  :::   ::: ::
Db    471 AANNDAGAMDHWTYWMGNNTMWMRRAWKMMNMAWCRRAYCNNNNNRACVWHKHKMWRWT  412

Qy   1965 GCCTCTAGGACAAAATTGACGTGTTCCGGGATCCTGCTGAATTAAGAACAACCCTAGGTG  2024
          |   :|::|:::    ::|::::   ::::|:||::::
Db    411 WKYMWKAACNNNNBKAMYMRVAWMYSRDTTNTDWMWTSDWBWHWYTVDYTMMRAWNNN  352

Qy   2025 CACCTGTCCCGATAGAGTCCCACCTGGTAGGCATTCATAGGGATTCGTGTATTTCCTGC  2084
            ::::::  :: :::|::|     ::::||::
Db    351 NNNNWRBCKTTSWMWMDHNNTHCYGNNTWGSAYBMAAMSMWAAGASNBVTYNWCWRMT  292

Qy   2085 AAAAAAAGCGATTAAGCTGGCTTCTAAAACTGGCTAGGCCGGATTCTGTGGCCTTCACTAC  2144
          :|::   ::::||::: |||:    :|:  :||
Db    291 YMGKTMTNNNNNNNKAWYYRTKTVAWCNNRYYDTAVWTBKRNYKYCYAYBWYBMYMGKH  232

Qy   2145 CAGGTGATTTTCATGTGATCCGTGCATTCTCAGCACTTTGCTATGTAACCCAAACTTAAGT  2204
          :|   :|:  :|:::     ||: :::::|:||:
Db    231 HWBWRRABHRSWNMWWVKCRNKYMVSWHYHAMRYBKWABAVGCNNNWKDRMAHHHWCAT  172

Qy   2205 CGACAACTATAAATATGCTACTTGCAGGATGTTATCACGACACAACTCCTAATCTACGGA  2264
             ::|:::::   :::||::::  :::::::
Db    171 NNNNMWWYAYMHMHKKGKAAWTNNKTABRDDHBAHVKTYWYWRYDYWCAMCMWNAKAK  112

Qy   2265 AGCCTAAGTTAGTTTTGCTCGGAGACAAGCAATTG  2300
          :: |||::  ::|:::     ::::   |:::::
Db    111 VRTAMKHMWYTDRYVSANNTGVRWMMWCMWWYS  76
```

```
      ATTORNEY/AGENT INFORMATION:
        NAME: McMasters, David D.
        REGISTRATION NUMBER: 33,963
        REFERENCE/DOCKET NUMBER: 240052.419C1
      TELECOMMUNICATION INFORMATION:
        TELEPHONE: (206) 622-4900
        TELEFAX: (206) 682-6031
  INFORMATION FOR SEQ ID NO: 209:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 51259 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
    SEQUENCE DESCRIPTION: SEQ ID NO: 209:
US-09-618-166-209

Query Match           2.0%;  Score 53.4;  DB 4;  Length 51259;
Best Local Similarity 54.3%;  Pred. No. 0.0011;
Matches  108;  Conservative  0;  Mismatches  91;  Indels  0;  Gaps  0;

Qy     314 CGCCCTGGCAGCGAGCTCCTGTAGTATCACCTGGTCGCCTCCAGCTCATGCTCGCAAGC 373
           ||||  || ||||||                            |||||||||||||||
Db     107 CTCCTTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTACAGCTCCTTCCACAT 166

Qy     374 CTCCAGGGCGGCCCCGGCCAGTGCTCCAACACTTCCACCTTTTACTCTTTCTTTCTTGG 433
           ||| |||||||||||||||||||||||||||                   ||||||||
Db     167 CTGCTCCTCCTCCTCCTGCTCCTCCTCCTCCTC 226

Qy     434 GCAGTCGTGCTCCGCACGACACCTTCTCCACCTTTTACTCTTTCTTTCTTTCTTTCTTGG 493
           ||||||||||||                       |||||||||||||||||
Db     227 CTCCTCCTCCTCCTCCTCCTGCTCCTCCTC 286

Qy     494 CCATCTTGGTATTTCA 512
           ||  ||| ||||| ||
Db     287 ACATTCTGTGTTTTACTCA 305

RESULT 5
US-09-424-283-6
; Sequence 6, Application US/09424283
; Patent No. 6437219
; GENERAL INFORMATION:
;   APPLICANT: Grimes, et al.
;   TITLE OF INVENTION: Sucrose binding proteins
;   FILE REFERENCE: 4630-50206
;   CURRENT APPLICATION NUMBER: US/09/424,283
;   CURRENT FILING DATE: 1999-11-19
;   PRIOR APPLICATION NUMBER: PCT/US98/10465
;   PRIOR FILING DATE: 1998-05-21
;   PRIOR APPLICATION NUMBER: US 60/047,568
;   PRIOR FILING DATE: 1997-05-22
;   NUMBER OF SEQ ID NOS: 15
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 3718
;   TYPE: DNA
;   ORGANISM: Glycine max
US-09-424-283-6

Query Match           1.9%;  Score 51.8;  DB 3;  Length 3718;
Best Local Similarity 96.4%;  Pred. No. 0.00063;
Matches   53;  Conservative  0;  Mismatches  2;  Indels  0;  Gaps  0;

Qy       7 TAATACGACTCACTATAGGGACGACGCGTGGTCGACGGCCCGGCGTCGCCACA 61
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db      99 TAATACGACTCACTATAGGGACGACGCGTGGTCGACGGCCCGGCGTCGCTGAGAAA 153

RESULT 6
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
```

```
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: PatentIn Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/781,891
      FILING DATE: 27-DEC-1996
      CLASSIFICATION: 800
  ATTORNEY/AGENT INFORMATION:
      NAME: No. 609020tenburg Ph.D., Carol
      REGISTRATION NUMBER: 39,317
      REFERENCE/DOCKET NUMBER: 240052.419
  TELECOMMUNICATION INFORMATION:
      TELEPHONE: (206) 622-4900
      TELEFAX: (206) 682-6031
  INFORMATION FOR SEQ ID NO: 209:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 51259 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
US-08-781-891-209

Query Match           2.0%;  Score 53.4;  DB 3;  Length 51259;
Best Local Similarity 54.3%;  Pred. No. 0.0011;
Matches  108;  Conservative  0;  Mismatches  91;  Indels  0;  Gaps  0;

Qy     314 CGCCCTGGCAGCGAGCTCCTGTAGTATCACCTGGTCGCCTCCAGCTCATGCTCGCAAGC 373
           ||||  || ||||||                            |||||||||||||||
Db     107 CTCCTTCTCCTCCTCCTCCTCCTCCTCCTCCTCCTACAGCTCCTTCCACAT 166

Qy     374 CTCCAGGGCGGCCCCGGCCAGTGCTCCAACACTTCCACCTTTTACTCTTTCTTTCTTGG 433
           ||| |||||||||||||||||||||||||||                   ||||||||
Db     167 CTGCTCCTCCTCCTCCTGCTCCTCCTCCTCCTC 226

Qy     434 GCAGTCGTGCTCCGCACGACACCTTCTCCACCTTTTACTCTTTCTTTCTTTCTTTCTTGG 493
           ||||||||||||                       |||||||||||||||||
Db     227 CTCCTCCTCCTCCTCCTCCTGCTCCTCCTC 286

Qy     494 CCATCTTGGTATTTCA 512
           ||  ||| ||||| ||
Db     287 ACATTCTGTGTTTTACTCA 305

RESULT 4
US-09-618-166-209
; Sequence 209, Application US/09618166
; Patent No. 6583112
; GENERAL INFORMATION:
;   APPLICANT: Fu, Ying-Hui
;              Yu, Chang-En
;              Oshima, Junko
;              Mulligan, John T.
;              Schellenberg, Gerald D.
;   TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
;      WERNER'S SYNDROME
;   NUMBER OF SEQUENCES: 209
;   CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Seed Intellectual Property Law Group
;      STREET: 701 Fifth Avenue, Suite 6300
;      CITY: Seattle
;      STATE: Washington
;      COUNTRY: USA
;      ZIP: 98104-7092
;   COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/09/618,166
;      FILING DATE: 17-Jul-2000
;      CLASSIFICATION: <Unknown>
```

```
;   APPLICANT: DORNER, F.
;   APPLICANT: SCHEIFLINGER, F.
;   APPLICANT: FALKNER, F. G.
;   TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
;   NUMBER OF SEQUENCES: 52
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Foley & Lardner
;     STREET: 1800 Diagonal Road, Suite 500
;     CITY: Alexandria
;     STATE: VA
;     COUNTRY: USA
;     ZIP: 22313-0299
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/232,463
;     FILING DATE:
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US/07/935,313
;     FILING DATE:
;     APPLICATION NUMBER: EP 91 114 300.6
;     FILING DATE: 26-AUG-1991
;   ATTORNEY/AGENT INFORMATION:
;     NAME: BENT, Stephen A.
;     REGISTRATION NUMBER: 29,768
;     REFERENCE/DOCKET NUMBER: 30472/114 IMMU
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (703)836-9300
;     TELEFAX: (703)683-4109
;     TELEX: 899149
;   INFORMATION FOR SEQ ID NO: 14:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 7218 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   IMMEDIATE SOURCE:
;     CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match         1.9%;  Score 50.8;  DB 1;  Length 7218;
Best Local Similarity   6.0%;  Pred. No. 0.0019;
Matches   25;  Conservative  217;  Mismatches  174;  Indels    0;  Gaps     0;

Qy  1276 TTTGAGCATGTTTATATTTAGGAGGACGTTTACCTGTTGAGGTAAATCTGAACGCTAAT 1335
              |   || ||      |||   |   || || :  ||  :: : ::
Db  1472 TCTATGCAAGTAGTTAAAGAGATAGAAGATTTCGTACRCRRRRRRRRRRRRRRRRRRRR 1413

Qy  1336 AAATCGGCTAAGCAAAATAATTTATCACCTGTTGATTCTAACAATTTGATGATGGACAAT 1395
                                     :: :: :: : :::::::::::::::::::::::::::
Db  1412 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1353

Qy  1396 ATTGATGAGGTGACTGACAAATGATTGAAGGCTTTAAAGGAGAGGATTGAGAAGGATAAATCT 1455
              :: :::::::: :::::::    ::::::::    ::::::::::::: :
Db  1352 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1293

Qy  1456 ACAATAAAAATGTAAAGAAGAGAAAGCATTCAAAGTGTGAGATCTCGGTGTGGAAGACTATTT 1515
              : :: ::::: :  ::::::::  :::: :::::   : :  :: : ::
Db  1292 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1233

Qy  1516 TGCCTCTTGGGGGTAAAAGACAACAAGTTTAGTAAGTTAGTAGTGGCCTCAAAATTGGGAGGGGCCCA 1575
              || :: : :::::: ::: :: :: : ::  :: :::  ::
Db  1232 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1173

Qy  1576 TGCAAGATTGTTAAAGTAATTGTTTTGGATTGACGGAGGCATTTCAAGGTGATCATCTAC 1635
              ::: : :::::: :: : :: :: :::::::: : :::: ::
Db  1172 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1113

Qy  1636 CTAGAGCTCTCAATGGGAGGAGGTGCTCGAAGACAATATTACCCATGTGTATGGCAAGAT 1691
```

```
Db  1112 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRATCGCCAAGCT 1057
               :::::: :          ::::::::    :::::   ||  ||||||

RESULT 7
US-09-249-585A-2/c
; Sequence 2, Application US/09249585A
; Patent No. 641702
; GENERAL INFORMATION:
;   APPLICANT: Horlick, Robert
;   TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
;   FILE REFERENCE: 0867/0D905
;   CURRENT APPLICATION NUMBER: US/09/249,585A
;   CURRENT FILING DATE: 1999-02-11
;   NUMBER OF SEQ ID NOS: 18
;   SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
;   LENGTH: 1926
;   TYPE: DNA
;   ORGANISM: Epstein Barr Virus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1926)
;   OTHER INFORMATION: coding strand of EBNA-1 DNA
US-09-249-585A-2

Query Match         1.8%;  Score 49.2;  DB 3;  Length 1926;
Best Local Similarity   57.0%;  Pred. No. 0.0024;
Matches   90;  Conservative   0;  Mismatches   68;  Indels    0;  Gaps     0;

Qy   329 CTCCTGTAGTATCACCTGCGTGCGCTCCAGCTCATGCTCGGCAAGCCTCCAGGGGCGGCCCG 388
             |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   466 CTCCTGCCCCTCCTCCTGCTCCTGCCCCCTCCTCCTGCTCCTCCTGCCCCTCCTGCCCCTCCTG 407

Qy   389 GCAGTGCTCCAACACTTTGCCTCCTCCTCCACAGCTCCTTCCACATGCAGTCGTGCTCCGC 448
             |||||||||  ||||| |||||||||||  |||||||||||||||||||||||||||||
Db   406 CCCCTCCTCCTGCTCCTGCCCCTCCTCCTCCTCCTGCCCCTCCTGCCCCCTCCTGCCCCTC 347

Qy   449 ACGCACCTTCTCCACCTTTTACTCTTTCTTTCTT 486
             |||||||||||||||||| || ||||||||||
Db   346 CTGCCCCTCCTCCTGCCCCTCCTCCTGCCCCT 309

RESULT 8
US-09-410-399-3/c
; Sequence 3, Application US/09410399
; Patent No. 6482587
; GENERAL INFORMATION:
;   APPLICANT: Robertson, Erle S.
;   APPLICANT: Cotter, Murray A.
;   TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
;   TITLE OF INVENTION:       to Genomic Host DNA
;   FILE REFERENCE: UM-03778
;   CURRENT APPLICATION NUMBER: US/09/410,399
;   CURRENT FILING DATE: 1999-10-01
;   NUMBER OF SEQ ID NOS: 6
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 1926
;   TYPE: DNA
;   ORGANISM: Epstein-Barr virus
US-09-410-399-3

Query Match         1.8%;  Score 49.2;  DB 4;  Length 1926;
Best Local Similarity   57.0%;  Pred. No. 0.0024;
Matches   90;  Conservative   0;  Mismatches   68;  Indels    0;  Gaps     0;

Qy   329 CTCCTGTAGTATCACCTGCGTGCGCTCCAGCTCATGCTCGGCAAGCCTCCAGGGGCGGCCCG 388
             |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   466 CTCCTGCCCCTCCTCCTGCTCCTGCCCCCTCCTCCTGCTCCTCCTGCCCCTCCTGCCCCTCCTG 407

Qy   389 GCAGTGCTCCAACACTTTGCCTCCTCCTCCTCCACAGCTCCTTCCACATGCAGTCGTGCTCCGC 448
             |||||||||  ||||| |||||||||||  ||||||||||||||||||||||||||||||
```

Db   406 CCCCTCCTCCTGCCCCTCCTGCCCCTCCTGCGCCTCCTGCCCCTCCTGCCCCTC 347

Qy   449 ACGCACCTTCTCCACCTTTTTACTCTTTTCTTTCTT 486
         ||||| ||| |||| || |||||
Db   346 CTGCCCCTCCTCTGCTCCTGCCCCTCCTGCTCCT 309

RESULT 9
US-09-050-863-2/c
; Sequence 2, Application US/09050863
; Patent No. 6114111
; GENERAL INFORMATION:
;   APPLICANT: Lao, Ying
;   APPLICANT: Hiang, Betty
;   APPLICANT: Payan, Don
;   TITLE OF INVENTION: Mammalian Protein Interaction Cloning
;   TITLE OF INVENTION: System
;   NUMBER OF SEQUENCES: 5
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;   STREET: 4 Embarcadero Center, Suite 3400
;   CITY: San Francisco
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 94111-4187
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/050,863
;   FILING DATE: 30-MAR-1998
;   CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Silva, Robin M.
;   REGISTRATION NUMBER: 38,304
;   REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (415) 781-1989
;   TELEFAX: (415) 949-8711
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2580 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: unknown
;   TOPOLOGY: unknown
;   MOLECULE TYPE: DNA
US-09-050-863-2

Query Match   1.8%; Score 49.2; DB 3; Length 2580;
Best Local Similarity   57.0%; Pred. No. 0.0029;
Matches   90; Conservative   0; Mismatches 68; Indels   0; Gaps   0;

Qy   329 CTCCTGTAGTATCACCTGGCTGCGCTCCAGCTCATGCTCCAAGCTCCAGGGGGGCCCG 388
         ||||| ||||||||||| ||||| | ||| ||| |||||
Db   849 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTG 790

Qy   389 GCAGTGCTCCAACACTTTCGCCTCCTCCTACAGCTCCTTCCACATGCAGTCGTGCTCCGC 448
         |||| ||||||| || || ||||||| |||
Db   789 CCCCTCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 730

Qy   449 ACGCACCTTCTCCACCTTTTTACTCTTTTCTTTCTT 486
         ||| ||| |||| || |||||
Db   729 CTGCCCCTCCTCTGCTCCTGCCCCTCCTGCTCCT 692

RESULT 10
US-09-359-081-2/c
; Sequence 2, Application US/09359081
; Patent No. 6316223
; GENERAL INFORMATION:
;   APPLICANT: Lao, Ying
;   APPLICANT: Hiang, Betty
;   APPLICANT: Payan, Don
;   TITLE OF INVENTION: Mammalian Protein Interaction Cloning
;   System
;   NUMBER OF SEQUENCES: 5
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;   STREET: 4 Embarcadero Center, Suite 3400
;   CITY: San Francisco
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 94111-4187
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/359,081
;   FILING DATE: 22-Jul-1999
;   CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 09/050,863
;   FILING DATE: <Unknown>
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Silva, Robin M.
;   REGISTRATION NUMBER: 38,304
;   REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (415) 781-1989
;   TELEFAX: (415) 949-8711
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2580 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: unknown
;   TOPOLOGY: unknown
;   MOLECULE TYPE: DNA
;   SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-359-081-2

Query Match   1.8%; Score 49.2; DB 3; Length 2580;
Best Local Similarity   57.0%; Pred. No. 0.0029;
Matches   90; Conservative   0; Mismatches 68; Indels   0; Gaps   0;

Qy   329 CTCCTGTAGTATCACCTGGCTGCGCTCCAGCTCATGCTCGCAAGCCTCCAGGGGGGCCCG 388
         ||||| ||||||||||| ||||| | ||| ||| |||||
Db   849 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTG 790

Qy   389 GCAGTGCTCCAACACTTTCGCCTCCTCCTACAGCTCCTTCCACATGCAGTCGTGTCCGC 448
         |||| ||||||| || || ||||||| |||
Db   789 CCCCTCCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCTCCTGCCCCTCCTGCCCCTC 730

Qy   449 ACGCACCTTCTCCACCTTTTTACTCTTTTCTTTCTT 486
         ||| ||| |||| || |||||
Db   729 CTGCCCCTCCTCTGCTCCTGCCCCTCCTGCTCCT 692

RESULT 11
US-09-130-114-1
; Sequence 1, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
;   APPLICANT: Horlick, Robert A.
;   APPLICANT: Damaj, Bassam B.
;   APPLICANT: Robbins, Alan K.
;   TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
;   TITLE OF INVENTION: From Multiple Transfected Episomes
;   FILE REFERENCE: 0867/1D903US1
;   CURRENT APPLICATION NUMBER: US/09/130,114
;   CURRENT FILING DATE: 1998-08-06
;   NUMBER OF SEQ ID NOS: 36
;   SOFTWARE: FastSEQ for Windows Version 3.0

```
; SEQ ID NO 1
;   LENGTH: 5452
;   TYPE: DNA
;   ORGANISM: VEBNA
US-09-130-114-1

Query Match     1.8%;  Score 49.2;  DB 2;  Length 5452;
Best Local Similarity  57.0%;  Pred. No. 0.0046;
Matches  90;  Conservative  0;  Mismatches  68;  Indels  0;  Gaps  0;

Qy    329 CTCCTGTAGTATCACCTGGCTCGCCTCCAGCTCATGCTGCCAAGCCTCCAGGGGCGGCCCG 388
              |||||  |     |   ||||||||||||||||||||||||||||||
Db   1956 CTCCTGCCCCTCCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCCTCCTG 2015

Qy    389 GCAGTGCTCCAACACTTCGGCTCCTTCCTACAGCTCCTTCCACATGCAGTCGTGCTCCGC 448
              | ||||     |||||||||||||||||||||||||||||||
Db   2016 CCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCCTCCTGCCCCCTC 2075

Qy    449 ACGCACCTTCTCCACCTTTTTACTCTTTCTTTCTCTT 486
              ||||  |||     |||||
Db   2076 CTGCCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCT 2113
```

RESULT 12
US-09-647-344A-14
```
; Sequence 14, Application US/09647344A
; Patent No. 6586180
; GENERAL INFORMATION:
;   APPLICANT: Ruffner, Duane E.
;   APPLICANT: Pierce, Michael L.
;   APPLICANT: Chen, Zhidong
;   TITLE OF INVENTION: Directed Antisense Libraries
;   FILE REFERENCE: T6678.PCT US
;   CURRENT APPLICATION NUMBER: US/09/647,344A
;   CURRENT FILING DATE: 2000-12-04
;   PRIOR APPLICATION NUMBER: PCT/US99/06742
;   PRIOR FILING DATE: 1999-03-28
;   NUMBER OF SEQ ID NOS: 50
; SEQ ID NO 14
;   LENGTH: 8705
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: pShuttle
US-09-647-344A-14

Query Match     1.8%;  Score 49.2;  DB 4;  Length 8705;
Best Local Similarity  57.0%;  Pred. No. 0.0062;
Matches  90;  Conservative  0;  Mismatches  68;  Indels  0;  Gaps  0;

Qy    329 CTCCTGTAGTATCACCTGGCTCGCCTCCAGCTCATGCTGCCAAGCCTCCAGGGGCGGCCCG 388
              |||||  |     |   ||||||||||||||||||||||||||||||
Db   7821 CTCCTGCCCCTCCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCCTCCTG 7880

Qy    389 GCAGTGCTCCAACACTTCGGCTCCTTCCTACAGCTCCTTCCACATGCAGTCGTGCTCCGC 448
              | ||||     |||||||||||||||||||||||||||||||
Db   7881 CCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCCTCCTGCCCCCTC 7940

Qy    449 ACGCACCTTCTCCACCTTTTTACTCTTTCTTTCTCTT 486
              ||||  |||     |||||
Db   7941 CTGCCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCT 7978
```

RESULT 13
US-08-910-647-1/c
```
; Sequence 1, Application US/08910647
; Patent No. 6251433
; GENERAL INFORMATION:
;   APPLICANT: Zuckermann et al.
;   TITLE OF INVENTION: Compositions and Methods for
;   TITLE OF INVENTION: Polynucleotide Delivery
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Chiron Corporation
;   STREET: 4560 Horton Street
;   CITY: Emeryville
;   STATE: California
;   COUNTRY: U.S.A.
;   ZIP: 94608-2916
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/910,647
;   FILING DATE:
;   CLASSIFICATION: 514
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Fujita, Sharon M.
;   REGISTRATION NUMBER: 38,459
;   REFERENCE/DOCKET NUMBER: 1218.002
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (510) 923-2706
;   TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 9600 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-08-910-647-1

Query Match     1.8%;  Score 49.2;  DB 3;  Length 9600;
Best Local Similarity  57.0%;  Pred. No. 0.0066;
Matches  90;  Conservative  0;  Mismatches  68;  Indels  0;  Gaps  0;

Qy    329 CTCCTGTAGTATCACCTGGCTCGCCTCCAGCTCATGCTGCCAAGCCTCCAGGGGCGGCCCG 388
              |||||  |     |   ||||||||||||||||||||||||||||||
Db    895 CTCCTGCCCCTCCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCCTCCTG 836

Qy    389 GCAGTGCTCCAACACTTCGGCTCCTTCCTACAGCTCCTTCCACATGCAGTCGTGCTCCGC 448
              | ||||     |||||||||||||||||||||||||||||||
Db    835 CCCCTCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCCTCCTGCCCCCTC 776

Qy    449 ACGCACCTTCTCCACCTTTTTACTCTTTCTTTCTCTT 486
              ||||  |||     |||||
Db    775 CTGCCCCCTCCTCCTGCTCCTGCCCCTCCTCCTGCTCCT 738
```

RESULT 14
US-09-620-925-1/c
```
; Sequence 1, Application US/09620925
; Patent No. 6468986
; GENERAL INFORMATION:
;   APPLICANT: Zuckermann et al.
;   TITLE OF INVENTION: Compositions and Methods for
;                       Polynucleotide Delivery
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRSS:
;   ADDRESSEE: Chiron Corporation
;   STREET: 4560 Horton Street
;   CITY: Emeryville
;   STATE: California
;   COUNTRY: U.S.A.
;   ZIP: 94608-2916
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/620,925
;   FILING DATE: 21-Jul-2000
;   CLASSIFICATION: <Unknown>
```

SEQUENCE CHARACTERISTICS:
    LENGTH: 10596 bases
    TYPE: NUCLEIC ACID
    STRANDEDNESS: single
    TOPOLOGY: linear
US-07-884-811-15

Query Match   1.8%;   Score 49.2; DB 1; Length 10596;
Best Local Similarity 57.0%; Pred. No. 0.0071;
Matches 90; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```
Qy   329 CTCCTGTAGTATCACCTGGCTCGCCTCCAGCTCATGCTCGCAAGCCTCCAGGGCGGCCCG 388
         |||||| ||    ||  ||  ||||||||| |||||||||||||||||||||
Db  2429 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCCTCCTG 2370

Qy   389 GCAGTGCTCCAACACTTTCGCCTCCTCCTACAGCTCCTTCCACATGCAGTCGTGCTCCGC 448
         || |||||||  |||  |||||||||||||||| |||||||||||||||||||||
Db  2369 CCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCCTCCTGCCCCCTCCTC 2310

Qy   449 ACGCACCTTCTCCACCTTTTTACTCTTTTCTTTCTT 486
         || ||||||||||  ||||||||||||||||||||
Db  2309 CTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCCTCCTGCTCCT 2272
```

Search completed: May 9, 2005, 02:22:57
Job time : 453 secs

PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 08/910,647
    FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
    NAME: Fujita, Sharon M.
    REGISTRATION NUMBER: 38,459
    REFERENCE/DOCKET NUMBER: 1218.002
TELECOMMUNICATION INFORMATION:
    TELEPHONE: (510) 923-2706
    TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 1:
    SEQUENCE CHARACTERISTICS:
        LENGTH: 9600 base pairs
        TYPE: nucleic acid
        STRANDEDNESS: single
        TOPOLOGY: linear
    MOLECULE TYPE: DNA (genomic)
    SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-620-925-1

Query Match   1.8%;   Score 49.2; DB 3; Length 9600;
Best Local Similarity 57.0%; Pred. No. 0.0066;
Matches 90; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```
Qy   329 CTCCTGTAGTATCACCTGGCTCGCCTCCAGCTCATGCTCGCAAGCCTCCAGGGCGGCCCG 388
         |||||| ||    ||  ||  ||||||||| |||||||||||||||||||||
Db   895 CTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCCTCCTG 836

Qy   389 GCAGTGCTCCAACACTTTCGCCTCCTCCTACAGCTCCTTCCACATGCAGTCGTGCTCCGC 448
         || |||||||  |||  |||||||||||||||| |||||||||||||||||||||
Db   835 CCCCTCCTGCTCCTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCCTCCTGCCCCCTCCTC 776

Qy   449 ACGCACCTTCTCCACCTTTTTACTCTTTTCTTTCTT 486
         || ||||||||||  ||||||||||||||||||||
Db   775 CTGCCCCTCCTGCTCCTGCCCCTCCTGCCCCCTCCTGCTCCT 738
```

RESULT 15
US-07-884-811-15/c
Sequence 15, Application US/07884811
Patent No. 5316921
GENERAL INFORMATION:
    APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
    TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
    NUMBER OF SEQUENCES: 21
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: Genentech, Inc.
        STREET: 460 Point San Bruno Blvd
        CITY: South San Francisco
        STATE: California
        COUNTRY: USA
        ZIP: 94080
    COMPUTER READABLE FORM:
        MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: patin (Genentech)
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/07/884,811
        FILING DATE: 19920518
        CLASSIFICATION: 530
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER:
        FILING DATE:
    ATTORNEY/AGENT INFORMATION:
        NAME: Dreger, Ginger R.
        REGISTRATION NUMBER: 33,055
        REFERENCE/DOCKET NUMBER: 755.1
    TELECOMMUNICATION INFORMATION:
        TELEPHONE: 415/225-3216
        TELEFAX: 415/952-9981
        TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 15:

```
              GenCore version 5.1.6
        Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:    May 8, 2005, 21:38:04 ; Search time 8593 Seconds
                                    (without alignments)
                                    11915.848 Million cell updates/sec

Title:          US-10-751-550-1
Perfect score:  2690
Sequence:       1 tctagataatacgactcact.........ccataactagtacaccatgg 2690
Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0
Searched:       34239544 seqs, 19032134700 residues
Total number of hits satisfying chosen parameters:   68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       EST:*
                 1: gb_est1:*
                 2: gb_est2:*
                 3: gb_htc:*
                 4: gb_est3:*
                 5: gb_est4:*
                 6: gb_est5:*
                 7: gb_est6:*
                 8: gb_gss1:*
                 9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	235	8.7	670	9	CL922874	CL922874 OA_ABa002
c 2	226.4	8.4	629	9	CL914408	CL914408 OA_ABa001
c 3	218.8	8.1	1020	9	CL947599	CL947599 OsIFSB005
c 4	212.4	7.9	1293	9	CL982445	CL982445 OsIPSC047
c 5	210.6	7.8	732	9	CL936467	CL936467 OA_ABa004
c 6	209	7.8	729	9	CL935937	CL935937 OA_ABa000
c 7	207.8	7.7	718	9	CL907647	CL907647 OA_ABa000
c 8	203.2	7.6	757	9	CL927201	CL927201 OA_ABa003
c 9	202	7.5	718	9	CL910742	CL910742 OA_ABa003
c 10	194	7.2	4152	9	CL958665	CL958665 OsIFCC001
c 11	183.4	6.8	669	9	CL916614	CL916614 OA_ABa001
c 12	176	6.5	707	9	CL922867	CL922867 OA_ABa002
c 13	175.8	6.5	678	9	CL908388	CL908388 OA_ABa000
c 14	175	6.5	902	8	BZ797683	BZ797683 PUGBP40TD
c 15	175	6.5	949	8	BZ797681	BZ797681 PUGBP40TB
c 16	172	6.4	757	9	CL918926	CL918926 OA_ABa002
c 17	171.6	6.4	722	9	CL149354	CL149354 107_330_1
c 18	166.8	6.2	813	9	CL907539	CL907539 OA_ABa000
c 19	164.4	6.2	850	9	CL941236	CL941236 OA_ABa005
c 20	164.2	6.1	779	8	BZ747430	BZ747430 PUDCK34TD
c 21	164.2	6.1	836	8	CC404666	CC404666 PUHJK33TD
c 22	163.2	6.1	660	9	CL923944	CL923944 OA_ABa002
c 23	162	6.0	909	8	CC436809	CC436809 PUHGQ26TD
c 24	161.4	6.0	801	9	CL903983	CL903983 OA_ABa000
c 25	161	6.0	749	9	CG867026	CG867026 ZMMBbb035
c 26	160	5.9	729	9	CC828183	CC828183 ZMMBbb017
c 27	160	5.9	773	9	CL917811	CL917811 OA_ABa002
c 28	159.8	5.9	650	8	CC374857	CC374857 PUEDJ53TD
c 29	159.4	5.9	722	8	CC008227	CC008227 PUDCY10TD
c 30	158.8	5.9	884	9	CG247085	CG247085 OG0AT79TH
c 31	158.4	5.9	697	9	CL933353	CL933353 OA_ABa004
c 32	158.2	5.9	966	8	BZ819943	BZ819943 PUGDI31TD
c 33	157.4	5.8	686	9	CL938390	CL938390 OA_ABa005
c 34	157.2	5.8	770	9	CL941728	CL941728 OA_ABa005
c 35	156.8	5.8	711	9	CL922324	CL922324 OA_ABa002
c 36	156.8	5.8	1028	9	CC969919	CC969919 ZUAB238TV
c 37	156.2	5.8	726	8	AQ274269	AQ274269 nbxb0034K
c 38	156	5.8	681	9	CL933514	CL933514 OA_ABa004
c 39	155.6	5.8	790	9	CC734662	CC734662 OGVAX46TH
c 40	155.6	5.8	838	9	CG002979	CG002979 ZUAD61TH
c 41	155.6	5.8	842	8	BZ795811	BZ795811 PUFDG84TB
c 42	155.6	5.8	927	8	CC347166	CC347166 OGOBH58TV
c 43	155.4	5.8	627	9	CC937241	CC937241 ZMMBbb022
c 44	155.2	5.8	775	9	CL906326	CL906326 OA_ABa000
c 45	154.6	5.7	655	8	AQ577400	AQ577400 nbxb0091A

ALIGNMENTS

```
RESULT 1
CL922874/c
LOCUS       CL922874           670 bp    DNA    linear    GSS 14-SEP-2004
DEFINITION  OA_ABa0028B23.r OA_ABa Oryza australiensis genomic clone
            OA_ABa0028B23 3', genomic survey sequence.
ACCESSION   CL922874
VERSION     CL922874.1  GI:5203943
KEYWORDS    GSS.
SOURCE      Oryza australiensis
ORGANISM    Oryza australiensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1 (bases 1 to 670)
AUTHORS     Kim,H.,Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
TITLE       OMAP Project
JOURNAL     Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR Primers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0028  row: B  column: 23
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..670
                     /organism="Oryza australiensis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4532"
                     /clone="OA_ABa0028B23"
                     /tissue_type="young leaves"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OA_ABa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

Query Match           8.7%;   Score 235;   DB 9;   Length 670;
Best Local Similarity 68.0%;  Pred. No. 2.2e-52;
Matches 387;  Conservative 0;  Mismatches 175;  Indels 7;  Gaps 4;
```

```
     source          1..629
                     /organism="Oryza australiensis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4532"
                     /clone="OA_ABa0016D07"
                     /tissue_type="young leaves"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OA_ABa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

Query Match       8.4%;    Score 226.4; DB 9; Length 629;
Best Local Similarity   69.9%;  Pred. No. 4.9e-50;
Matches  362;  Conservative   0;  Mismatches 151;  Indels   5;  Gaps   4;

Qy   506 ATTTTCACAAATGTCCCCTACACAAATGATAAATCACCAAAACTACATGGAGCTTGCTAGTT 565
             |||||  ||  ||  || ||  ||||||  |||||  |||||||  ||||||||||||||
Db   527 ATTTTGATATATATTCACCTGCATACAAATAATCACCAACCCTCATGGAAAGAGTTAGTA 468

Qy  .566 ATAAACTCAATTCTAAGTTGGTGTTATTGAGTGGATTTGTGTGAAAGTTGGTGGT 625
             |  ||||||  ||||||||  |||||  ||||||||  ||  ||||||||||
Db   467 AATAACCTACCGCGTCACGTTGATGTGCATCGTTGTTCAATTTATGCAGGGATTGACGGT 408

Qy   626 TAGAAATAGG-AGTTAAGGACCGCCAACAAGATCCCCACACTTAGCCCTTGCTCATCC 684
             |  |||||   |||||  |||  |||  ||  ||||  |||||||  |||||||||
Db   407 CAGATTTCGGCAGTTAAGGACCGTCAACAAGCTCCCCACACTTTGCCCTTTGCTCGTCC 348

Qy   685 TCGAGTAAAGTTCAAG-GACTAAGGTGGACATCTTCCTCAAATGGTACGATGCCTGCATA 743
             ||||||||||||||   ||||||||||||| ||||||||  ||||  |||||||||
Db   347 TCGAGTAAAGGTCGAGATACTAAGGTGGATCATGATAAATTTGATACAATGCTAACATA 288

Qy   744 TAAGTTATTCCAAGCCTCACCTATACATGTGAACTTTGAAGTGTCTACCACGGCCATCTTG 803
             |||||||||||||||||||  |||||  ||||||||  |||||||||||||||||||||
Db   287 CAAGTTATTCCTAGACCTATCATATATTGAATCTGGATGTGTCTACCATGATATCATA 228

Qy   804 GGTGGTTGAGAAATGGAACAGAGATCAGAATCCAGTCATCTTACCTCTCTTGCTAGATAA 863
             |||||||||||||||||  |||  ||  |||||||||  |||||||||||||||  |||
Db   227 AGTTATTGAAAAATGGAATAGTCTAGGTGTAAGTCACTTTTATTCTTCTGCTTCAATGA 168

Qy   864 CTTGGG--TTTTGTAAGGTTTCAAATTAAAACATAGTCTTGCTCCTCAAATGATTCT 921
             |||||   ||||| ||||||||||||  |||||  || |||||||||||||||||||
Db   167 CTTGGGGTTCTTAAGAGGTTTTCAAAAGAAAAACATAACTTTACTTTACTCCTGTGGAATGT 108

Qy   922 CTCATATAGCTCAATGTGTATGGGTTTTCCACCAAGGCAATGTTTTGCTCTTCTTTCATCCT 981
             |||||  |||||||||||||  || ||||||||||||||||||||  ||  |||||||||||
Db   107 CTCATGTCGTCAATGTGTTTGAATCCTCCACCAAGGCAAT-TATTGCCTAATTTTTCTCCT 49

Qy   982 ACTTCTAATATTTCTTTTGTGGAGCTTAGGGTAGGGAA 1019
             |||||||||||  |||||||||||||||||||||||||
Db    48 ACTTCTAATAAGGCTTATGTGGAGCTTCAAGGTAGGGAA 11


RESULT 3
CL947599
```

```
Qy   517 TGTCCCCTACAAATGATAAATCACCAAAACTCATGGAGCTTGCTAGTTATAAAACTCTAA 576
             ||||||  ||||||||||  ||||||||  |||||||||||||||||  ||||  |||
Db   588 TATTCGCATGCATACAAATAATCACCAACCCTCGTGGAAAGAGTTAGTAAATAACCCTAC 529

Qy   577 TTCTAAGTTTGGTGTTATTGAGTGGATTTGTGTGGTGGTTAGAAATAGG- 635
             ||||||||  |||||||||||  ||||||  |  |||||  ||||||||
Db   528 GGCTACGTTGATATGCATCGTTGTCAATTTTGCAGGAATTGATGGTTGGATTCTGGC 469

Qy   636 AGTTAAGGACCGCCAACAAGATCCCCACACTTAGCCCTTTGCTCATCCTTCGAGTAAAGT 695
             |||||||||||  ||||||||||||||  ||||||||||||||||  |||||||||||
Db   468 AGTTAAGGACCATCAACAAGCTCCCCACACTTAGCCCTTTGCTCTTCTTCGAGTAAAGG 409

Qy   696 TCAAGGACTAAGGTGGAACATCTTCCTCAAATGGTACGATGCCTGCATATAAGTTATTCCA 755
             ||||||  ||||||||| ||||  ||||  ||||||  |||||||||||  ||||||||
Db   408 TTGAATACTAAGGTGGGTTATGAATTGTTTTGGTACAATGCTAACATACACAAGTTATTCTA 349

Qy   756 AGCCTCACCTATACATGTGAACTTTGAAGTGTCTACCACGGCCATCTTGGGTGGTTGAGAA 815
             ||||||||  |||||| ||||||||| | |||||||||||||||||||||||||||||||
Db   348 AGACTTACCCGTATCTGTGAATCTGGATGTGTCTACCATGATATCCTGAGCTATTGTGGA 289

Qy   816 ATGGAACAGATCAGAATCCAGTCATCTTCTTCCTTAGATAACTTGGGTTTTG 875
             |||||||||||||||||  ||||||| ||  ||  ||||  |||| |||||
Db   288 ATGGAACAGTCTAAGTTTAAGTTATGTTTATTCCTTCGCTTAAATGACTTGGGGGGTTTT 229

Qy   876 TAAGTTTTC---AAATTAAAACATAGTCTGTCCTCAAATGATTCTCTCATATAGC 931
             |  |||||   |||||||||  ||  ||  ||||||||  ||||||||||||||
Db   228 CAAAAGTTTTCCAAAAGAAAAACATAACTTTACTCCTCCTGATGATTCTCCATGTGTCAC 169

Qy   932 TCAATGTGTATGGTTTCTCACCAAGGCCAATGTTTTGCCTCTTTTCATCCTCCTAATA 991
             ||||||||| ||||||  |||||||||  |||||||||  |||||||||||  ||||
Db   168 TCAATATGTTGATTCCTCCACCAAGGC-ATTCCTCACCTGCCTTAATTCCTCCTACTCTAATA 110

Qy   992 TTTCTTTTGTGGAGCTTAGGGGTAGGGAATGAAAAGGAAGCATACTTGCATTGCATATGTT 1051
             |||||||||||||||||||  ||||||||||||  ||||||||||||||||||||||||
Db   109 AAGGTTTATGTGGAGCTTCAAGGTAGGGAA--AATCATGAGGCATATTTGCATCAAATATTTT 51

Qy  1052 ACTAAGTCAAAAACCAAATCTGAGAGAA 1080
             |||||||||||  ||||  |||||||||
Db    50 GTAAAGTCAAACAGTAGATCCAAGAAGAA 22


RESULT 2
CL914408/c
```

RESULT 4

```
LOCUS       CL982445     1293 bp    DNA    linear   GSS 21-SEP-2004
DEFINITION  OsIFSC047866 Oryza sativa Express Library Oryza sativa (indica
            cultivar-group) genomic, genomic survey sequence.
ACCESSION   CL982445
VERSION     CL982445.1  GI:52419372
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
  ORGANISM  Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1 (bases 1 to 1293)
  AUTHORS   Ma,L., Wang,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
```

Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.

```
FEATURES             Location/Qualifiers
     source          1..1020
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
```

ORIGIN

```
Query Match             8.1%;  Score 218.8;  DB 9;  Length 1020;
Best Local Similarity  70.9%;  Pred. No. 6.6e-48;
Matches 343;  Conservative  0;  Mismatches 117;  Indels 24;  Gaps 3;

Qy  1232 GAGCTCATCATTATATAAGCATGGTGATACCAAAATTACTCCTTTTGAGCATGTTTATA 1291
              |||||||||  ||||| |||||||||| |||| |||||||||||||||||| |||| |
Db   533 GGGCTCATCGTATATCTAGACATGGTGCTACTAAAGTTACTCCTTTTGAGCTTGTGTATG 592

Qy  1292 TTTAGGAGACGTTTTACCTGTTGAGGTAAATCTGAACGCTAATAAATCGGCTAAGCAAA 1351
             || ||||| ||||| || |||||||||||||| ||||| ||| | |||| ||||  |||
Db   593 GTCAAGAGGGCCGTTTTGCCCGTTGAGGTAAATCTGGACGCGTTTATAGATTGGCTAAACAAA 652

Qy  1352 ATAATTTATCACCTGTTGATTCTAACAATTTGATGATGACAATATTGATGAGGTGACTG 1411
             || |   | || |||||||||  | |||||||||||  | ||||||||||||||| | |
Db   653 ATGACCTTGATGCTGTTGATTACCACAATTTGATGTTGGCAATATTGATGAGGTAACCG 712

Qy  1412 ACAAATGATTGAAGGCTTTAAAGAGGAT-------AAATCT 1455
             |||| ||||||||  |||| | |||||              |  |||
Db   713 ACAAACGATTGAAAGCTTTGAAGGAGGATTGAGAAAGATAAGCTTCGGGTTGCTAAAGCTT 772

Qy  1456 ACAATAAAAATGTAAAGAAGAAAGCATTCAAAGTGTGAGATCTGGTGTGTGGAAGACTATTT 1515
             |||| | |||   || |||  |  ||||| |||| ||||| | |||  |||||||| | ||
Db   773 ACAACAGAAAGTACGAGGAAAATATTTTCAGATTGTGAGTTGTTGTTTGGGAAGACGATCT 832

Qy  1516 TGCCTCTTGGGGGTAAAGACAACAAGTTAGTAAGTGGCTGCCTCCAAAATTGGGAGGGCCCA 1575
             ||||| ||||   | |||| |||||||| | ||| |||  | || |||| |||||||| ||
Db   833 TGCCTATTGGGACGAAAAG-CAACAAGTTTGGTAAATGGTCACCAAGTTTGGGAAGGACCC 891

Qy  1576 TGCAAGATTGTTAAAGTAATTGTTTTGGAT------TGACGGAGGCATTTCAAGGTGAT 1628
             | | | | |||| ||| | |||||| ||||          || || ||| | || || |
Db   892 TATAGAGTGTTAAAGTCATTTTTGGGAATTCTTACATGCTAGAGACGCGCTGAAGGGAGAT 951

Qy  1629 CATCTCACCTAGAGCTCTCAATGGGAGGTGCTCGAAGACATATTACCCATGTGTATGGCAA 1688
             |||  |  | ||| | || |||||||   | |||||| || || || |||| | ||||||
Db   952 CATTTGCCTAAAGTCATCAACGGAGATACTTGAAGAAATTCTATCCAAGTGTTTGGCAA 1011

Qy  1689 GATG 1692
             |||
Db  1012 GACG 1015
```

RESULT 5

```
LOCUS       CL936467     732 bp    DNA    linear   GSS 14-SEP-2004
DEFINITION  OA_ABa0047D02.r OA_ABa Oryza australiensis genomic clone
            OA_ABa0047D02 3', genomic survey sequence.
ACCESSION   CL936467
VERSION     CL936467.1  GI:52067031
KEYWORDS    GSS.
SOURCE      Oryza australiensis
```

Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
Wong,G.K.S., Deng,X.W. and Wang,J.

```
  TITLE     An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
  JOURNAL   Unpublished (2004)
  COMMENT   Contact: Chen Chen
            Department of Bioinformatic
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES             Location/Qualifiers
     source          1..1293
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
```

ORIGIN

```
Query Match             7.9%;  Score 212.4;  DB 9;  Length 1293;
Best Local Similarity  70.0%;  Pred. No. 4e-46;
Matches 339;  Conservative  0;  Mismatches 121;  Indels 24;  Gaps 3;

Qy  1232 GAGCTCATCATTATATAAGCATGGTGATACCAAAATTACTCCTTTTGAGCATGTTTATA 1291
              |||||||||  ||||| |||||||||| |||| |||||||||||||||||| |||| |
Db   806 GGGCTCATCGTATATCTAGACATGGTGCTACTAAAGTTACTCCTTTTGAGCTTGTGTATG 865

Qy  1292 TTTAGGAGACGTTTTACCTGTTGAGGTAAATCTGAACGCTAATAAATCGGCTAAGCAAA 1351
             || ||||| ||||| || |||||||||||||| ||||| ||| | |||| ||||  |||
Db   866 GTCAAGAGGGCCGTTTTGCCCGTTGAGGTAAATCTGGACGCGTTTATAGATTGGCTAAACAAA 925

Qy  1352 ATAATTTATCACCTGTTGATTCTAACAATTTGATGATGACAATATTGATGAGGTGACTG 1411
             || |   | || |||||||||  | |||||||||||  | ||||||||||||||| | |
Db   926 ATGACCTTGATGCTGTTGATTACCACAATTTGATGTTGGCAATATTGATGAGGTAACCG 985

Qy  1412 ACAAATGATTGAAGGCTTTAAAGAGGAT-------AAATCT 1455
             |||| ||||||||  |||| | |||||              |  |||
Db   986 ACAAACGATTGAAAGCTTTGAAGGAGGATTGAGAAAGATAAGCTTCGGGTTGCTAAAGCTT 1045

Qy  1456 ACAATAAAAATGTAAAGAAGAAAGCATTCAAAGTGTGAGATCTGGTGTGTGGAAGACTATTT 1515
             |||| | |||   || |||  |  ||||| |||| ||||| | |||  |||||||| | ||
Db  1046 ACAACAGAAAGTACGAGGAAAATATTTTCAGATTGTGAGTTGTTGTTTGGGAAGACGATCT 1105

Qy  1516 TGCCTCTTGGGGGTAAAGACAACAAGTTAGTAAGTGGCTCCAAAATTGGGAGGGCCCA 1575
             ||||| ||||   | |||| |||||||| | ||| |||  || |||| |||||||| ||
Db  1106 TGCCTATTGGGACGAAAAG-CAACAAGTTTGGTAAATGGTCACCAAGTTTGGGAAGACCC 1164

Qy  1576 TGCAAGATTGTTAAAGTAATTGTTTTGGAT------TGACGGAGGCATTTCAAGGTGAT 1628
             | | | | |||| ||| | |||||| ||||          || || ||| | || || |
Db  1165 TATAGAGTAGTCCAAGTCATTTTCCGGAATTCTTACATGCTAGAGACGCGCTGAAGGGAGAT 1224

Qy  1629 CATCTCACCTAGAGCTCTCAATGGGAGGTGCTCGAAGACATATTACCCATGTGTATGGCAA 1688
             |||  |  | ||| | || |||||||   | |||||| || || || |||| | ||||||
Db  1225 CATTTGCCTAAAGTTATCAATGGGACATATTTGAAGAAAATTCTATCCAAGTGTTTGGCAA 1284

Qy  1689 GATG 1692
             |||
Db  1285 GATG 1288
```

First record

```
ORGANISM   Oryza australiensis
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
           Ehrhartoideae; Oryzeae; Oryza.
REFERENCE  1 (bases 1 to 732)
AUTHORS    Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
           Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
TITLE      OMAP Project
JOURNAL    Unpublished (2004)
COMMENT    Contact: Rod A. Wing
           Arizona Genomics Institute
           University of Arizona
           Forbes Building Room 303, Tucson, AZ 85721-0036, USA
           Tel: 520 626 9595
           Fax: 520 621 1259
           Email: http://genome.arizona.edu
           PCR PRimers
           FORWARD: TAA TAC GAC TCA CTA TAG GG
           BACKWARD: CAC TCA TTA GGC ACC CCA
           Plate: 0047  row: D  column: 02
           Seq primer: CAC TCA TTA GGC ACC CCA
           Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..732
                     /organism="Oryza australiensis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4532"
                     /clone="OA_ABa0047D02"
                     /tissue_type="young leaves"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OA_ABa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN
Query Match           7.8%;    Score 210.6;  DB 9;  Length 732;
Best Local Similarity 73.2%;   Pred. No. 1.1e-45;
Matches 311;  Conservative  0;  Mismatches 109;  Indels  5;  Gaps  3;

Qy   659 CCCCACACTTAGCCCTTGCTCATCCTCGAGTAAGTTCAAGGACTAAGGTGGAACATCT 718
             ||||||||||||||||| |||||||||||||||||| |||||||||||||||||||||
Db     1 CCCCACACTTAGTCCTTTGCTCGTCCTCGAGTAAGGTTAAGTACTAAGGTGGATCATGA 60

Qy   719 CCTCAAATGGTACGATGCCTGCATATAAGTTATTCCAAGCCTCACCTATACATGTGAACT 778
             || ||| ||||  |||||||||||||||||||||| |||||||||||||| |||||
Db    61 CTTCTTTGGTACAATGTTGACATACAAGTTATTCCAAAACATACCTATATCTGTGAATT 120

Qy   779 TTGAAGTGTCTACCACGCCATCTTGGGTGTTGAGAAATGGAACAGATCCAGAATCCAGTC 838
             ||||||||||| |||||||||||||||||||||||| ||||||||||||||
Db   121 TGGATGTGTCTACCATGTTATCCTGAGTTATTGAAGAATGGAACAGTCTAGGTTTAAGTT 180

Qy   839 ATCTTTACCTCTCTGCTTAGATAACTTGGGTTTTGTAAGGTTTTCAAATTT---AAAA 895
             ||||||||||||||| || |||||| |||| |||||
Db   181 ACTTTAATTCCTTCTGCTTAAATGACTTGAGGTTTTCAATAAGTTTTCAAAATGAAAAAA 240

Qy   896 CATAGTCTTGCTCCTCCAAATGATTCTCATATAGCTCAAATGTGTATGGTTTCTCACCAA 955
             ||||||||| || || |||||||||  ||||||||||||||| ||||||||||||||||
Db   241 CATAGCTTTTACTCCTCCAGGATGATTCTTCTAGGTCACTCAAATGTGTTTGATTCCTCACCAA 300

Qy   956 GGCAATGGTTTTGCCTCTTTTCATCCTACTTCTAATATTTCTTTTGTGGAGCTTAGGGTAG 1015
             ||||||||||||||||||||| ||||||||| |||||||||||||
Db   301 GGCATTGATTTGCCTCTTTTC--TCCTACTTCTTTATAAGGCTTATGTGGGGCTCAAGATAG 359

Qy  1016 GGAATGAAAAGGAAGCATACTTGCATTGCATATGTTACTAAGTCAAAACCAAATCTGAG 1075
             |||||||||||||| |||||| || || ||||||||||||
Db   360 GGAA-AATCATGAGGCATACTTGCATCAAATATTTTGTAAGTCAAAAAGTATATCCAAG 418

Qy  1076 GAGAA 1080
             |||||
Db   419 AAGAA 423

RESULT 6
CL935937/c
```

Second record

```
LOCUS      CL935937                729 bp    DNA    linear   GSS 14-SEP-2004
DEFINITION OA_ABa0046H01.r OA_ABa Oryza australiensis genomic clone
           OA_ABa0046H01 3', genomic survey sequence.
ACCESSION  CL935937
VERSION    CL935937.1  GI:52065973
KEYWORDS   GSS.
SOURCE     Oryza australiensis
ORGANISM   Oryza australiensis
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
           Ehrhartoideae; Oryzeae; Oryza.
REFERENCE  1 (bases 1 to 729)
AUTHORS    Kim.H., Yu,Y., Stum,D., Yost,D., Rao.K., Luo,M., Jetty,R.,
           Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
TITLE      OMAP Project
JOURNAL    Unpublished (2004)
COMMENT    Contact: Rod A. Wing
           Arizona Genomics Institute
           University of Arizona
           Forbes Building Room 303, Tucson, AZ 85721-0036, USA
           Tel: 520 626 9595
           Fax: 520 621 1259
           Email: http://genome.arizona.edu
           PCR PRimers
           FORWARD: TAA TAC GAC TCA CTA TAG GG
           BACKWARD: CAC TCA TTA GGC ACC CCA
           Plate: 0046  row: H  column: 01
           Seq primer: CAC TCA TTA GGC ACC CCA
           Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..729
                     /organism="Oryza australiensis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4532"
                     /clone="OA_ABa0046H01"
                     /tissue_type="young leaves"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OA_ABa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN
Query Match           7.8%;    Score 209;  DB 9;  Length 729;
Best Local Similarity 65.1%;   Pred. No. 2.9e-45;
Matches 372;  Conservative  0;  Mismatches 190;  Indels  9;  Gaps  4;

Qy   517 TGTCCCCTACAAATGATAAATCACCAAAACTCATGGAGCTTGCTAGTTATAAACTCTAA 576
             | |||||||||||| |||| ||||||||||||||
Db   590 TATTCACCTGCATACAAATAATCACCAACACTCATCGGAAAGAGTTAGTAAATAACCCTAC 531

Qy   577 TTCTAAGTTTGGTGTTTATTGAGTGGATTTGTGTGAAAGTTCGGTGGTGTAGAAATAGG- 635
             ||||||||||||||||||||| || |||||||||||||||
Db   530 CGCTATGTTGATGTGCATCGTTGTCAGTTTATGCAGGGAGGCGGGGGATTTTGGC 471

Qy   636 AGTTAAGGACCGCCAACAAGATCCCCACAGACTTAGCCCTTGCTCATCCTCGAGTAAAGT 695
             ||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db   470 AGTTAAGGACCGTCAACAAGCTCCCCTACACACATAGCCATTGCTGTCCTTGAGTAGAAGG 411

Qy   696 TCAAGGACTAAGGTGGAACATCTCCTCAAATGCTCGATGCCTGCATATAAGTTATTCCA 755
             ||||||||||||||||||||||| ||||||| || |||||||||| |||||||
Db   410 TTGGAATACTAAGGTGGATCATGATCATGTACAATGCTAATATACAAATTGTTCCT 351

Qy   756 AGCCTCACCTATACATGTGAACTTTGAAGTGTCTACCACGCCATCTTGGGTGGTTGAGAA 815
             |||||||| |||||||| || ||||||||| |||||||||||||
Db   350 AGACTTACCTGTATCGTGAATCTGAATTTGTCTACCATTGTATCTTAAGTTATTGAAGA 291

Qy   816 ATGGAACAGATCAGAATCCAGTCATCATCTTTACCTCTTGCTTAGATAACTTGGGTTTT- 874
             ||| ||||||||| || |||| || ||||||||  ||||||||||||||
Db   290 AGGGAACAATCTAGGTTTAAGTTACTTTTATTTCTTCTGCTTAAATGACTTGGGGTTTTC 231

Qy   875 -----GTAAGGTTTTCAAATTTAAAACATAGTCTTGCTCCTCCAAATGATTCTCATATA 929
                  || ||||||||||||
Db   230 AAGAAGTTCTTTTAGAAAAAGAAAATATAATTTTACTTCCTCGATGATTCTCATATC 171
```

```
Qy   930 GCTCAATGTGTATGGTTTCTCACCAAGGCAATGTTTTGCCTCTTTTCATCCTACTTCTAA 989
         ||||||||||||| || ||||||  |||||||| |||| |||||||||||| ||| |||||||||
Db   170 ACTCAATGTGTTTGATTCCCCACCAAGGTAATG-CTTGCCTAATTTTCTCCTACTTCTAA 112

Qy   990 TATTTCTTTTGTGGAGCTTAGGGTAGGGAATGAAAAGAAGCATACTTGCATTGCATATG 1049
         |||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   111 TAAGGCTTATGTGGAGCTCAAGGTAGGGAA-AATCAAAAGGCATACTTGCATCAAATATT 53

Qy  1050 TTACTAAGTCAAAAACCAAATCTGAGGAGAA 1080
         || |||||||| | |||  ||||||||||
Db    52 TTGTAAAGTCAAACAGAAGATCCAAGAAGAA 22

RESULT 7
CL907647/c
LOCUS       CL907647               718 bp    DNA        linear    GSS 14-SEP-2004
DEFINITION  OA_ABa0006L01.r OA_ABa Oryza australiensis genomic clone
            OA_ABa0006L01 3', genomic survey sequence.
ACCESSION   CL907647
VERSION     CL907647.1  GI:52016526
KEYWORDS    GSS.
SOURCE      Oryza australiensis
  ORGANISM  Oryza australiensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 718)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0006  row: L  column: 01
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..718
                     /organism="Oryza australiensis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4532"
                     /clone="OA_ABa0006L01"
                     /tissue_type="young leaves"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OA_ABa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

Query Match           7.7%;  Score 207.8;  DB 9;  Length 718;
Best Local Similarity 66.5%;  Pred. No. 6.1e-45;
Matches  377;  Conservative   0;  Mismatches  177;  Indels  13;  Gaps   5;

Qy   515 AATGTCCCCTACAAATGATAAATCACCAAAACTCATGGAGCTTGCTAGTTATAAACTCT 574
         || |||||| ||||||||||||| |||| |||| ||||||  |||
Db   577 ATTATTCACCTTCATACAAATATTCACCAACACTTGTGGAAGGGATTAGTAAATAACCAT 518

Qy   575 AATTCTAAGTTTGGTGTTTATTGAGTGGATTTGTGTGAAAGTTGGTGGTTAGA-AATA 633
         ||||| |||| ||  || |||| || ||||| ||||| |||
Db   517 ACCACTATGTTGATGTTCATCTTTTATTCATTTTATGCAGGCTTTGATGCCATATTTTA 458

Qy   634 GGAGTTAAGGACCCGCCACCAAGATCCCCACCTTAGCCCTTTGCTCATCCTCGAGTAAA 693
         ||||||||| |||| |||||||| ||||| ||||||||||| ||||||||
Db   457 GTAGTTAAGGACCACCACCAAGCTCCTCCACACTTAGTCCTTGCTCGTCCTGGTGAGTAAA 398

Qy   694 GTTCAAGGACTAAGGTGGAACATCTCCTCAAATGTACGATGCCTGCATATAAGTTATTC 753
```

Right column:

```
Db   397 GGTTAAGTACTAAGGTGGATAATGATTTCTTCTGGTACAATGCTGACATACAAGTTATTC 338

Qy   754 CAAGCCTCACCTATACATGTGAACTTTGAAGTGTCACCAGCCATCTTGGGTGGTTGAG 813
         ||  |||||||||||| || |||||||||||| || ||||||||||||| |||||||
Db   337 CAAAACATACTATATCTGTGAATCTGGATGTGTCTATCATCTTATCCTGGGTTATTGAA 278

Qy   814 AAATGGAACAGATCAGAATCCAGTCATCTTTACCTCTCTTGCTTAGTAACTTGGGTTT 873
         |||||||||||||||||||||||||| |||||||||||||||| ||||||||
Db   277 GAATGGAATAGTCTAGGTTGAGTTATCTTCCTTCCTCGCTTAAGACACTTGGGTTTT 218

Qy   874 TGTAAGGTTTTCAAATTTAAAACATAGTCTTGCTCCTCCAAATGATTCTCTCATATAGCTC 933
         ||||| |||||| |  | |||||||||||||||||||||||||||||| ||||||||||
Db   217 CAAATGGA--------AAAACATAGCTTTACTCCTCGGATGATTCTCTCATGTTAATC 167

Qy   934 AATGTGTATGGTTTCTCACCAAGGCAATGTTTGCCTCTTTTCATCCTACTTCTAATATT 993
         |||||||||||||||||||||||||||||| ||||||||||||||||||  | |||||||
Db   166 -ATGTGTTTGATTCCTCACCAAGGCATTGTTTCACATGTTT-TTCCTATTTCTAATCAG 109

Qy   994 TCTTTTGTGGAGCTTAGGGTAGGGAATGAAAAGAAGCATACTTGCATTGCATATGTTAC 1053
         |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   108 GCTTATATGGGCTCAAGGTAGGGAA-AATCATGAGACATATTTTCATCAAATATTTGT 50

Qy  1054 TAAGTCAAAAACCAAATCTGAGGAGAA 1080
         ||||||||| | |||  |||||||||
Db    49 AAAGTCAAGAGTATATCCAAGAAGAA 23

RESULT 8
CL927201/c
LOCUS       CL927201               757 bp    DNA        linear    GSS 14-SEP-2004
DEFINITION  OA_ABa0034A06.f OA_ABa Oryza australiensis genomic clone
            OA_ABa0034A06 5', genomic survey sequence.
ACCESSION   CL927201
VERSION     CL927201.1  GI:52048628
KEYWORDS    GSS.
SOURCE      Oryza australiensis
  ORGANISM  Oryza australiensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 757)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0034  row: A  column: 06
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..757
                     /organism="Oryza australiensis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4532"
                     /clone="OA_ABa0034A06"
                     /tissue_type="young leaves"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OA_ABa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

Query Match           7.6%;  Score 203.2;  DB 9;  Length 757;
Best Local Similarity 65.9%;  Pred. No. 1.1e-43;
```

Matches 373; Conservative 0; Mismatches 183; Indels 10; Gaps 5;

```
Qy    519  TCCCCCTACAAATGATAAATCACCAAAACTCATGGAGCTTGCTAGTTATAAACTCTAATT  578
                |||| |||||||  || |||||||| |||| ||||||||| |||||||||||||||||
Db    581  TTCACCTGCATACAAATAATCACTAACACTCATGGAAAGTGTTAGTAAATAGCCCCACTA  522

Qy    579  CTAAGTTGGTGTTTATTTGAGTGGATTTGTGTGAAAGTGGTGTGTTAGAAATAGG-AG   637
           |||    |||||   |||||||||||||||||| |||||||||||||||||||| ||
Db    521  CTACATTGATGTGCATCGTTTGTTCAATTATGTGCAATTAGCAATTGACGGTCGGATTTTTGGTAG  462

Qy    638  TTAAGGACCGCCAACAAGATCCCCACCACTTAGCCCTTTGCTCATCCTCGAGTAAAGTTC  697
              |||| ||| ||||||| |||||||||||||||||||||||| ||||||||||||
Db    461  GTAAGGACCGTCAACAAGCTGCCCCACCACTTAGCCCTTTGTTGCCTTGAGTAAAGGTT  402

Qy    698  AAGGACTAAGGTGGAACATCTCCTCAAATGGTACGATCCTGCATATAAGTTATTCCAAG  757
           ||||||  |||| |||||||||||||| ||||||||||||||||||||||||||||||
Db    401  GAATACTAAGGTGGATCATGATTTCTTTGGTACAATGCTAACATACAAGTTATTCCAAG  342

Qy    758  CCTCACCTATACATGTGAACTTGAGTGTCTACCACGCCATCTTGGGTGGTCGTTGAGAAAT  817
           ||||||    |||| ||||||  ||||| |||||   |||||||   ||  ||
Db    341  ACTTACCCGTATCTGTTAATCTGGATTTGTCTACCATGATATAGTGAGTTATTGAAGAAT  282

Qy    818  GGAACAGATCAGAATCCAGTCATCTTTACCTCTCTTGCTTAGAGATAACTTGGGTTTTGTA  877
           |||||  ||  ||||||||| |||| |||| |||| |||| ||| |||||  ||||
Db    281  GGAACAATCTAGGTTTAAGTTACTTTTATTCCTCTGCTTAAATGACTTGGGGGTTTTCA   222

Qy    878  AGGTTTT---CAAATTTAAAACATAGTCTTGCTCCTCAAATGATTCTCTCATATAGCTCA  934
           |||||||   ||  ||||  ||| |||||||| ||||||||||||||| ||
Db    221  AAGGTTTTCAAAAAAAGAAAAACATAACTTTACTCCCTGATGATTCTCTCATGTCACTCA  162

Qy    935  ATGTGTATGGTTTCTCCACCAAGGCAATGTTTTGCCTCTTTTCATCCTCACTTCTAATATTT  994
           ||||||||||      ||||||||||  ||| ||| ||| ||||||||||||| |||||||
Db    161  ATGTGTTTGATT---ACCAAGGCCATTACTTTGTCTCTATTTTC--TCCTACTTCTAATAAAG  107

Qy    995  CTTTTGTGGAGCTTAGGGTAGGGGAATGAAAGGAAGCATACTTGCCATTGCCATATGTTACT  1054
           |||||||||||||||||||||||||||| |||  |||||||||||||||||||| |||
Db    106  CTTCTGTGGATCTCAAGGTAGGGAA-AATCATGTTACATACTTGCATCAAAATGGTGCGGA   48

Qy    1055 AAGTCAAAAACCAAATCTGAGGAGAA 1080
           ||||||||| |  || |||| || ||
Db    47   AAGTCAAAACAGTAGATGCAAGAAAAA  22
```

RESULT 9

```
CL910742/c
LOCUS         CL910742                          718 bp    DNA       linear   GSS 14-SEP-2004
DEFINITION    OA_ABa0011A10.f OA_ABa Oryza australiensis genomic clone
              OA_ABa0011A10 5', Genomic survey sequence.
ACCESSION     CL910742
VERSION       CL910742.1  GI:52019621
KEYWORDS      GSS.
SOURCE        Oryza australiensis
  ORGANISM    Oryza australiensis
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
              Ehrhartoideae; Oryzeae; Oryza.
REFERENCE     1  (bases 1 to 718)
  AUTHORS     Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
              Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE       OMAP Project
  JOURNAL     Unpublished (2004)
COMMENT       Contact: Rod A. Wing
              Arizona Genomics Institute
              University of Arizona
              Forbes Building Room 303, Tucson, AZ 85721-0036, USA
              Tel: 520 626 9595
              Fax: 520 621 1259
              Email: http://genome.arizona.edu
              PCR Primers
              FORWARD: TAA TAC GAC TCA CTA TAG GG
              BACKWARD: CAC TCA TTA GGC ACC CCA
              Plate: 0011 row: A  column: 10
              Seq primer: TAA TAC GAC TCA CTA TAG GG
```

```
FEATURES             Location/Qualifiers
     source          1..718
                     /organism="Oryza australiensis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4532"
                     /clone="OA_ABa0011A10"
                     /tissue_type="young leaves"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OA_ABa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN
```

Class: BAC ends.

```
Query Match       7.5%;    Score 202;  DB 9;   Length 718;
Best Local Similarity 72.1%;   Pred. No. 2.3e-43;
Matches 320;  Conservative  0; Mismatches 115;  Indels 9;  Gaps 4;
```

```
Qy    632  TAGGAGTTAAGGACCGCCAACAAGAATCCCCACACTTAGCCCTTTGCTCATCCTCGAGTA  691
                |||||||||||||||  ||   |||||||||||||||||||||||||||||||||
Db    440  TGGTAGTTAAGAACCATCAACAAGCTCCCCACCACTTAGCCCTTTGCTCATACTCGAGTA  381

Qy    692  AAGTTCAAGGACTAAGGTGGAACATCTCCTCAAATGGTACGAGTGCTGCATATAGTTAT   751
                || ||  |||| || |||||| |||| |    ||    |||||  |||||||||
Db    380  AACATTAAACACTATGGTGATCATTTTT----TGGTAAAATGTGACGTACAAGTTAT   325

Qy    752  TCCAAGCCTCACCTATACATGTGAACTTTGAAGTGTCTACCACGCCATCTTGGGTGGTTG  811
           ||||   ||| || |||| ||||||| |||  ||||| |||| |||||||| ||||||
Db    324  TCCAAAACATACCTGTATTGTGGATCTTGANTGTGTCTATCATGTGGGTTATCTGGGTTATTG  265

Qy    812  AGAAATGGAACAGATCAGA-ATCCAGTCATCTTTACCTCTCTTGCTTAGTAATAACTTGGGT  870
              |||||||||||||| || |||| ||| ||||| |  ||| |||| |||| |||||||
Db    264  AAGAATGGAACAGTTTAGATTTTAAGTTACCTTCATTCCTTCTGCTTAAGTGACTTGGGG   205

Qy    871  TTTTGTAAGGTTTTCAAAATT---AAAACAGTAGTCTTGCTCCCTCAAATGATTCTCTCATA  927
           |||||  ||||||||||||||   ||||||||||| |||| |||| ||||| |||||||
Db    204  TTTTCAATAGGTTTTCAAAATGGAAAAACATAGCTTACTTCCTCAGATCTCCTCCATG   145

Qy    928  TAGTCAAATGTATGTATGGTTTCTCACCAAGGCAATGTTTTGCCTCTTTCATCCTACTTCT  987
           |||||||||||| ||| |||||||||||||  |||||||||||||||| ||||||||||
Db    144  TCACTCAAATGTGTTTGATTCCTCCCCAAGGTTATTGCCTCACCTATTTTCTTCTTACTTAT   85

Qy    988  AATATTTCTTTTGTGTGGAGCTTAGGGTAGGGGAATGAAAAAGGAAGCATACTTGCATTGCATA  1047
               ||||||     |||| ||||||  ||| || ||  |||||| |||||||||||||
Db    84   AATAAGGTTTTATCTGGAGCTCAAGGTAGGGAA-AATCATGAGGCATACTTGCATCAAATA   26

Qy    1048 TGTTACTAAGTCAAAAACCAAATC  1071
           |||  ||||||||||||||||||
Db    25   TTTTGTAAAGTCAAAAAGGTAGATC  2
```

RESULT 10

```
CL958665
LOCUS         CL958665                          4152 bp   DNA       linear   GSS 21-SEP-2004
DEFINITION    OSIFCC001881 Oryza sativa Express Library Oryza sativa (indica
              cultivar-group) genomic, genomic survey sequence.
ACCESSION     CL958665
VERSION       CL958665.1  GI:52372080
KEYWORDS      GSS.
SOURCE        Oryza sativa (indica cultivar-group)
  ORGANISM    Oryza sativa (indica cultivar-group)
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
              Ehrhartoideae; Oryzeae; Oryza.
REFERENCE     1  (bases 1 to 4152)
  AUTHORS     Ma,L., Wangc,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
              Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
              Wong,G.K.S., Deng,X.W. and Wang,J.
  TITLE       An analysis of transcriptional regulation of the rice genome and
              its comparison to Arabidopsis
  JOURNAL     Unpublished (2004)
COMMENT       Contact: Chen Chen
              Department of Bioinformatic
              Beijing Institute of Genomics
```

JOURNAL
```
              Chinese Academy of Sciences, Beijing 101300, China
              Tel: 86-10-80481559
              Fax: 86-10-80488676
              Email: chenchen@genomics.org.cn
              Rice genomic sequence.
              Class: exon-trapped.
```

FEATURES   Location/Qualifiers
```
     source          1..4152
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:39946"
                     /clone_lib="Oryza sativa Express Library"
                     /note="Oryza sativa exon trapped genomic sequences "
```

ORIGIN

```
Query Match    7.2%;     Score 194;  DB 9;  Length 4152;
Best Local Similarity 67.4%;   Pred. No. 5.8e-41;
Matches 329;   Conservative   0;  Mismatches 135;  Indels  24;  Gaps  3;

Qy  1232 GAGCTCATCATTTATATAAGCATGGTGATACCAAAATTACTCCTTTGAGCATGTTTATA 1291
         ||||||||  |  |||||||||||||||     ||||||| |   |||| |||||||||
Db  3665 GGGCTCATCGTATATCTAAGCATGGTGCAACAAAGTAACACCATTTGAATTAGTATATG 3724

Qy  1292 TTTAGGAGGAGCGTTTACCTGTTGGAGGTAAATCTGAACGGTAAATCGGCTAAGCAAA 1351
         |||  |||||| |||| |||||| |||||| |||| ||   |||| ||||| |||||||
Db  3725 GGCAAGGAGCCATTTGCATGTTGAGGTGAATCTGGACGCCCTTAGATTGGCCAGACAAA 3784

Qy  1352 ATAATTTATCACCTGTTGATTCTAACAATTTGATGATGGACAATATTGATGAGGTGACTG 1411
         ||||||| || | || |  ||  || |||| |||||| |   ||  ||||||| ||| ||
Db  3785 ATAATTTGTCGGCAGTAGATTATCACAACTTAATGATGATGAGGAATAGATGAGGTTAGTG 3844

Qy  1412 ACAAATGATTGAAGGCTTTAAAGGAGGATTGAGAAGGATAAAT-----------CT 1455
         | |||||||||||  ||||   | ||| ||||||| |||||||              |
Db  3845 ATGAAAGATTGAAAGCTTTGAGAGAGATTGAGAAAGATAAATTGAGAGTAGCTAAAGCTT 3904

Qy  1456 ACAATAAAAATGTAAAGAAGAAAAGCATTCAAAGTGTGAGATCTCGGTGTGTGGAAGACTATTT 1515
         ||||||||| | | ||||| ||| |  |||||| | | |||| |  |||  | |||||  |||
Db  3905 ACAATAAAAAGGTGAAGAAAAAATCGTTTCAAATTGGAGATTTAGTGTCGGAAAACAATTC 3964

Qy  1516 TGCCTCTTGGGGTAAAAGACAACAAGTTTAGTAAGTGGCCTCAAAATTGGGAGGGCCCA 1575
         ||||| || ||| |||| ||  ||  | || || |||| |||||||||||| |||| ||
Db  3965 TGCCTGTT-GGTTCAAAAGATAATAAATTGGGAAATGTCTCCAAATTGGGAAGGGCCA 4023

Qy  1576 TGCAAGATTGTTAAAGTAAT-------TGTTTTGGATTGACGGAGGCATTCAAGGTGAT 1628
         |  |  ||||| ||| ||||          |||  ||  |  ||||| ||||| || |
Db  4024 TATAGAATTGTAGAAAATAGTGCCCCGGAATTCTTATTTTGTGCAAGTCTCACGAGGATAT 4083

Qy  1629 CATCTTACCTAGAGCTTCTCAATGGGAGGGTGCTCGAAGACATATTACCCATGTGTATGGCAA 1688
         || | ||| ||||||| | |||||||| | ||  |  | |||||| | |||||| | |||||
Db  4084 AAGTTGCCTAGAGCTCTTAATGGGAGATATTTAAAAAGGTACTATCCAAGTGTTTGGCAA 4143

Qy  1689 GATGTTTA 1696
         |||| |||
Db  4144 GATGCTTA 4151
```

RESULT 11
CL916614/c
```
LOCUS       CL916614                 669 bp    DNA     linear   GSS 14-SEP-2004
DEFINITION  OA_ABa0019F04.r OA_ABa Oryza australiensis genomic clone
            OA_ABa0019F04 3', genomic survey sequence.
ACCESSION   CL916614
VERSION     CL916614.1  GI:52027398
KEYWORDS    GSS.
SOURCE      Oryza australiensis
  ORGANISM  Oryza australiensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1 (bases 1 to 669)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
```

JOURNAL
COMMENT
```
            Unpublished (2004)
            Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR Primers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0019  row: F  column: 04
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
```

FEATURES   Location/Qualifiers
```
     source          1..669
                     /organism="Oryza australiensis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4532"
                     /clone="OA_ABa0019F04"
                     /tissue_type="young leaves"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OA_ABa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
```

ORIGIN

```
Query Match    6.8%;     Score 183.4;  DB 9;  Length 669;
Best Local Similarity 65.3%;   Pred. No. 2.8e-38;
Matches 331;   Conservative   0;  Mismatches 171;  Indels  5;  Gaps  4;

Qy  521 CCCCTACAAATGATAAATCACCAAAACTCATGGAGCTTGCTAGTTATAAACTCTAATTCT 580
        |||| | ||||| ||||| ||||  | || |||| |  | ||| || ||| |||| | |
Db  507 CACCTGCATACAAATATTAACCAACACTTGTGGAACAGATTAGTAGTAAATAACCTTATCGCT 448

Qy  581 AAGTTTGGTGTTTATTGAGTGGATTTTGTGTGAAAGTTGGTGGT-TAGAAATAGGAGTT 639
        | ||| | | | |||  ||| |||||  | | | | |||| ||  ||| ||| | |||
Db  447 ACATTGATGTTGATCTTTCTCCGTTTTATGCAGGGATTGATGGTCGAATTTTAGTATTT 388

Qy  640 AAGGACCGCCAACAAGATCCCCACCACTTAGCCCTTGCTCATCCTCGAGTAAAGTTCAA 699
        |||| | | ||||| |||||||||||||||| ||||||||  |||| |||||||| |||
Db  387 AAGGATCATCAACATGCTCCCCACCACTTAGTCCTTTGCTCGTCCTTCGGGTAAAGCTTAA 328

Qy  700 GGACTAAGGTGGAACATCTCCTCAAATGGTACGATGCCTGCATATAAGTTATTCCAAGCC 759
        | ||||||||||| ||| | | |||| ||||   ||| |||||| |||| |||||||| ||
Db  327 GTACTAAGGTGGATCATGACTTATTTGGTATAAATGCTGACATACAATGTATTCCAAACC 268

Qy  760 TCACCTATACATGTGAAC-TTTGAAGTGTCTACCACGCCATCTTGGTGGTTGAGAAATG 818
        | ||| | ||  |||||| |||||| |||||||||| |||||||||  | |||| ||||
Db  267 ATACCTGTACCCGTGAACTTTGAACGTGTCTACCACATGTCATCTTGGGTAATTGAAGAATG 208

Qy  819 GAACAGATCAGAATCCAGTCATCTTACCTCTCTTGCTTAGTGAATAACTTGGGTTTTTGTAA 878
        | ||||  | ||| | | | |||||  | || |||||  ||||||     ||| |||  ||
Db  207 GTACAGCCTTGATTCAAATCACCCCTACTTCTGATAAAT-ATCTTAGCGTTTTCCTTAT 149

Qy  879 GGTTTTCAAATTAAAACATAGTCTTGCTCCTCCAAATGATTCTCTCAATGATTCTCTCAATGT 938
        | | |  |||| ||||||||  |||  |  | |||| ||||||||||| ||||||||||||||
Db  148 GCTCATAGAAACATAACTTTACTTTTACTCCTCAATTGATTCTCTCAATTGCTCAATGT 89

Qy  939 GTATGGTTTCTCACCAAGGCAATGTTTGCCTCTTTTCATCCTAC--TTCTAATATTTCT 996
        ||  |  ||||||||||||| |||||||||| |||  ||||||    |||||||| | ||
Db  88 GTTTTATTCTCACCAAGGCCCTGATGTTGCCTATTCTTATTCCTACCTTTCTAATAAGTT 29

Qy  997 TTTGTGGAGCTTAGGGTAGGGAATGAA 1023
        | ||| |||||||||||||  |||| |
Db  28 TATGTGTAGCTTAGGGTAGAAAAATAA 2
```

RESULT 12
CL922867/c
```
LOCUS       CL922867                 707 bp    DNA     linear   GSS 14-SEP-2004
DEFINITION  OA_ABa0028B19.r OA_ABa Oryza australiensis genomic clone
            OA_ABa0028B19 3', genomic survey sequence.
ACCESSION   CL922867
```

```
VERSION     CL922867.1  GI:52039929
KEYWORDS    GSS.
SOURCE      Oryza australiensis
  ORGANISM  Oryza australiensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 707)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0028  row: B  column: 19
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..707
                     /organism="Oryza australiensis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4532"
                     /clone="OA_ABa0028B19"
                     /tissue type="young leaves"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OA_ABa"
                     /note="vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

Query Match              6.5%;  Score 176; DB 9; Length 707;
Best Local Similarity   65.3%;  Pred. No. 3e-36;
Matches 337;  Conservative    0; Mismatches 170;  Indels    9; Gaps    5;

Qy    521 CCCCTACAAATGATAAATCACCAAAACTCATGGAGCTTGCTAGTTATAAAACTCTAATTCT 580
Db    518 CACCTGCATACAAATATTCACCCAGCACTTGTGAACGAATTAGTAAATAACCCTACTGCT 459

Qy    581 AAGTTTGGTCGTTTATTTGAGTGGATTTGTTGTGAAAGTTGGTGGTTAGAAATAGG--AGT 638
Db    458 ATGTTGATGTTTATCTTTTATCATTTATGCAGGGATTGACAGTTCGGATTTTAGTACTT 399

Qy    639 TAAGGACCGCCAACAAGATCCCCCACACTTAGCCCTTTGCTCATCCTCGAGTAAAGTTCA 698
Db    398 CAAGGACTTTTAACAATCTCCCCCACACTTGCCCATTGCTCGTTCCTCGAGTAAAGGTTA 339

Qy    699 AGGATAAGGTGGAACATCTCCTCAAATGCTACGATGCCTGCATATAGTTATTCCAAGC 758
Db    338 AGTACTAAGGTGGATCATGACGTGTTTTGGTATATATGCTGACATACAACTTATAACAAA 279

Qy    759 CTCACCTATACATGTGAACTTGAAGTGTCTCACCACGCCATCTTGGGTGGTTGAGAAATG 818
Db    278 CATATCTGCACCTGTGAAACTTGATGTGTCTATTCATGTTATCTTGTGTGGATTGAAGAATG 219

Qy    819 GAACAGATCAGAATCCAGTCATCTTTACCTCTCTTTGCTAGATAAACTTGGGTTTTGTAA 878
Db    218 GAACAGTCTAGGGTTAAGTCATCTTCATTCCTTCGCTAAGTGACTCGGGTCATTCAATA 159

Qy    879 GGTTTTCAAATTTA--AAACATAGTCTTGCTCCTCAAATGATTCTCTCATATAGTCTCAA 935
Db    158 AATTTTTAAAACGGATAAAACATAACTTTACTCCTTGGA-GATTCTCATGTTCGCTCAA 100

Qy    936 TGTGTATGGTTTCTCACCAAGGCCAATG-TTTTGCCCTCTTTCATCCTAC--TTCTAATAT 992
Db     99 TGTGTTTGATTCCTCCTATTCTGCTAATTCTTATCCTACCTTTCTTTATAA 40
```

```
Qy    993 TTCTTTGTGGAGCTTAGGTAGGAGGAATGAAAAGGA 1028
Db     39 GGCTTATGTGGAGCTCAGGTAGGGAGGAAATAAGAGTA 4

RESULT 13
CL908388/c
LOCUS       CL908388               678 bp    DNA     linear    GSS 14-SEP-2004
DEFINITION  OA_ABa0007L13.r OA_ABa Oryza australiensis genomic clone
            OA_ABa0007L13 3', genomic survey sequence.
ACCESSION   CL908388
VERSION     CL908388.1  GI:52017267
KEYWORDS    GSS.
SOURCE      Oryza australiensis
  ORGANISM  Oryza australiensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 678)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0007  row: L  column: 13
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..678
                     /organism="Oryza australiensis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4532"
                     /clone="OA_ABa0007L13"
                     /tissue type="young leaves"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OA_ABa"
                     /note="vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

Query Match              6.5%;  Score 175.8; DB 9; Length 678;
Best Local Similarity   64.8%;  Pred. No. 3.3e-36;
Matches 370;  Conservative    0; Mismatches 192;  Indels    9; Gaps    7;

Qy    515 AATGTCCCCTACAAATGATAAATCACCAAAACTCATGGAGCTTGCTAGTTATAAAACTCT 574
Db    588 ATTATTCACCTGGATACAAATATTTACCACACTGTTGAAAGAGTTAATAAAATAACCCT 529

Qy    575 AATTCTAAGTTGGTCGTTTATTTGAGTGGATTTGTTGTGAAAGTTGGTGGTTAGA-AATA 633
Db    528 ACCACTACGTTGATGAGACATAGTTGTTCCATTTATGCAGGGATTGATGGTCGGATATTA 469

Qy    634 GGAGTTAAGGACCGCCAACAAGATCCCCCACACTTAGCCCTTTGCTCATCCTCGAGTAAA 693
Db    468 GTCGTTAAGGACCATAAACAAGCTCCCCCACACTTGCCCTTTGTCATCCTCGTGTAAA 409

Qy    694 GTTCAAG-GACTAAGGTGGAACATCTCCTCAAATGCTCGCATATAAGTTATT 752
Db    408 GGCCCGATACTAAGGTGGATCATGATAAATTTGGTGTAGAGCAAACATACAAGTTATT 349

Qy    753 CCAAGCGCTTCACCTATACATGTGAACTTGAAGTGTCTCACCACGCCATCTTGGGTGGTTGA 812
Db    348 CCTAGAACCTACTCATAACTGTGAATCTATGTGTCTACCATGTCTACCATGCGAATTCTAAGT-ATTGA 290

Qy    813 GAAATGGAACAGATCAGAATCCAGTCATCTTTACCTCTCTTCTTAGATAACTTGGGTTT 872
```

```
Db     289 AAAATGGAACAG-TCTAGATATAAGCTACTATATTCTGCTGCATTAAATAACTTGGGGTT 231
           ||||||||||||  ||  ||||||||||||||| |||  ||| |||||||| |||||||||
Qy     873 TTGTAAGGTTTC--AAAATTAAAACATAGTCTTGCTCCTCCAAATGATTCTCATATA 929

Db     230 TTGATAAGTTTCAAAAATGTAAAACATGACTTTACTCTGTGTGATGATTCTCATGTC 171
           ||||| |||||||| ||| |||||||||| ||||||| |||||| |||||||||
Qy     930 GCTCAATGTGTATGGTTCTCACCAAGGCAATGTTTTGCCTCTTTTCATCCTACTTCTAA 989

Db     170 ACTCGAGGTGTTGATTCCTCACCATGGCAAT-ACTTGGCCAAAAAATTTTCCTACTTCTAA 112
           ||||||||||||||  |||||||||||||| | ||||| |||||||||||||||||||||
Qy     990 TATTTCTTTTGTGGAGCTTAGGGTAGGGAATGAAAAGGAAGCATTGCATTGCATATG 1049

Db     111 TAAAGCTTATGTGCGAGCTCAAGGTAGGGAA-AATCATCGCGCATACTTGCATCAAATATT 53
           ||||||||||  ||||| ||||||||||||  |||||||||  |||||||||||||||
Qy    1050 TTACTAAGTCAAAAACCAAAATCTGAGGAGAA 1080

Db      52 TTTTAAAGTCAGATAAAAGATTCAAGAAGAA 22
           ||||||||||||||||||||||| ||||||
```

```
RESULT 14
BZ797683/c
LOCUS       BZ797683              902 bp    DNA     linear   GSS 17-MAR-2003
DEFINITION  PUGBP40TD ZM_0_6_1.0_KB Zea mays genomic clone ZMMBTa331G08,
            genomic survey sequence.
ACCESSION   BZ797683
VERSION     BZ797683.1  GI:29000291
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 902)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
            Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished (2003)
COMMENT     Other GSSs: PUGBP40TB
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1..902
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa331G08"
                     /clone_lib="ZM_0_6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
ORIGIN

Query Match             6.5%;  Score 175;  DB 8;  Length 902;
Best Local Similarity   64.7%;  Pred. No. 5.9e-36;
Matches  319;  Conservative   0;  Mismatches  150;  Indels  24;  Gaps   3;

Qy    1232 GAGCTCATCATTTATATAAGCATGTGATACCAAAATTACTCCTTTTGAGCATGTTATA 1291
Db     844 GGGCTCATGAGAATCAAAACTTCATGCCACTAAAGTTTCTCCTTTTGAGCTTATGTATG 785

Qy    1292 TTTAGGAGGACGTTTACCTGTTGAGGTAAATCTTGAGGTAAATCGGCTAAGCAA 1351
Db     784 GAAATGAAGGCCATCTTGCCTATAGAAATAAAATCTAACCACTATTAGGTTTGCCAGGCAAA 725

Qy    1352 ATAATTTATCACCTGTTGATTCTAACAATTGATGATGACAATATTGATGAGGTGACTG 1411
```

```
Db     724 ATAATTTAGATGTCGATGAGTACCATGATCATTTAATGATGACAATATTGACGAGGTGACCG 665
           ||||||||| |||| ||||| |||||| || |||||||||||||||||||| ||||||||||
Qy    1412 ACAAATGATTGAAGGCTTAAAGAGGATTGAGGAAGGATAAA------------TCT 1455

Db     664 ATAAGAGGTTGATGGCTTTGAAGGAAATTGAAAAGGACAAACTTATAGTCGCCAAGGTCT 605
           || || ||| ||| |||| ||| |||||||| |||| ||||   ||||||||| |||||||
Qy    1456 ACAATAAAAATGTAAAGGACAGAAAGCATTCAAAGTGTGAGATCTCGTTGTGGAAGACTATTT 1515

Db     604 ACAATAAGAGAGTTAAGACTAAAATCATTTCAGGTTGGAGATTTGGTGTGGGAGACAATGT 545
           |||||| || | || || |  || ||||| |||| |||||| || || ||| ||||||| |
Qy    1516 TGCCTCCTTGGGGGTAAAAGAACAACAAGTTTAGTAGTAGTGGCTCAAAAATTGGGAGGGCCCA 1575

Db     544 TGCC-ATTGAAGATTAAGGATTACAAGTTTGACAGTGGTCTCCAGTTTCCAAATATT 486
           |||| ||||||| ||||| || ||||||||||| |||||||||||||||| ||||
Qy    1576 TGCAAGATTGTTAAAGTAATTGTTTTGGAT-------TGACGGAGGCATTCAAGGTGAT 1628

Db     485 TATAAGGTTGTACATGTAATCCCTGGTAATGCTTATATGTTGGAAACATTACAAGGAAAT 426
           | || |||| |||| ||| ||| |||||  |  ||||| ||||||||| |||| ||||||
Qy    1629 CATCTACCTAGAGCTCTCAATGGAGGTGCTCGAAGACATATTACCCATGTGTATGCAA 1688

Db     425 AAGTTACATAAAGCATTGAATGGTCGTTTCCTAAAACAATACTATCCTAATACATGGTAA 366
           || || ||| |||| ||| ||||| ||| |||||||||||| ||||| |||||||| ||
Qy    1689 GATGTTTAGGCTAG 1701

Db     365 GATGTTTAGAAAG 353
           |||||||||   |
```

```
RESULT 15
BZ797681
LOCUS       BZ797681              949 bp    DNA     linear   GSS 17-MAR-2003
DEFINITION  PUGBP40TB ZM_0_6_1.0_KB Zea mays genomic clone ZMMBTa331G08,
            genomic survey sequence.
ACCESSION   BZ797681
VERSION     BZ797681.1  GI:29000287
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 949)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
            Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished (2003)
COMMENT     Other GSSs: PUGBP40TD
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1..949
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa331G08"
                     /clone_lib="ZM_0_6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
ORIGIN

Query Match             6.5%;  Score 175;  DB 8;  Length 949;
Best Local Similarity   64.7%;  Pred. No. 6e-36;
Matches  319;  Conservative   0;  Mismatches  150;  Indels  24;  Gaps   3;

Qy    1232 GAGCTCATCATTTATATAAGCATGTGATACCAAAATTACTCCTTTTGAGCATGTTATA 1291
```

Mon May 9 13:05:36 2005   us-10-751-550-1.rst   Page 10

```
Db         ||||||||  |||   ||   ||  |    ||   ||    | |||  ||   ||||||||||||   |  |||
           GGGCTCCATAGAATATCAAAACTTCATGCCACTAAAGTTTCTCCTTTTGAGCTTATGTATG 330
Qy    1292 TTTAGGAGGACGTTTTACCTGTTGAGGTAAATCTGAACGCTAATAAATCGGCTAAGCAAA 1351
           ||||||||  ||  ||   ||||||| ||||||| |  | |  | |||||||  | |||
Db     331 GAAATGAAGCCATCTTGCCTATAGAAATAAATCTAACCACTATTAGGTTTGCCAGGCAAA 390

Qy    1352 ATAATTTATCACCTGTTGATTCTAACAATTTGATGATGACAATATTGATGAGGTGACTG 1411
           ||||||  ||||||||  |||||||||||| ||||||||||||||  ||||||||||
Db     391 ATAATTTAGATGTCGATGAGTACCATGATTTAATGATGGACAATATTGACGAGGTGACCG 450

Qy    1412 ACAAATGATTGAAGGCTTTAAAGGCTTAAAGGAGATTGAGAAGGATAAA----------TCT 1455
           |||| ||||||| ||||| ||||||| |  ||||||||||||| |||          |||
Db     451 ATAAGAGGTTGATGGCTTTGAAGGAAATTGAAAAGGACAAACTTATAGTCGCCAAGGTCT 510

Qy    1456 ACAATAAAAATGTAAAGAAGAAAGCATTCAAAGTGTGAGATCTGGTGTGAAGACTATTT 1515
           |||||||  |||||| ||| |||| ||||| |||| ||||| |||||||||||| | ||
Db     511 ACAATAAGAAGTAAAGACTAAATCATTTCAGGTTGGAGATTTGGTGTGAAGACAATGT 570

Qy    1516 TGCCTCTTGGGGTAAAGACAACAAGTTTAGTAAGTGGCCTCAAAATTGGGAGGGCCCA 1575
           ||||  ||| |||||||  | |||||| |||||| |  |||  |||||| | | |||
Db     571 TGCC-ATTGAAGATTAAGGATTACAAGTTTGACAGTGGTCTCCCAAGTTTGAAAGGGACT 629

Qy    1576 TGCAAGATTGTTAAAGTAATTGTTTTGGAT------TGACGGAGGCATTCAAGGTGAT 1628
           | ||||||| || ||| |  || || ||||      |  | | |||  ||| ||||
Db     630 TATAAGGTTGTACATGTAATCCCTGGTAATGCTTATATGTTGGAAACATTACAAGGAAAT 689

Qy    1629 CATCTACCTAGAGCTCTCAATGGGAGGTGCTCGAAGACATATTACCCATGTGTATGGCAA 1688
           |||  |||| ||  |||  ||||| || | |||||||||| ||||| |||| |||||||
Db     690 AAGTTACATAAAGCATTGAATGGTCGTTTCGTTTCCTAAAAACAATACTATCCTAAATACACATGGTAA 749

Qy    1689 GATGTTTAGCTAG 1701
           |||||||| ||
Db     750 GATGTTTAGAAAG 762


Search completed: May  9, 2005, 04:46:22
Job time : 8601 secs
```